

GenCore version 4.5
• Copyright (c) 1993 - 2000 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2000, 09:14:06 ; Search time 768.4 seconds
(without alignments)
12051.847 Million cell updates/sec

Title: US-09-068-377-2
Perfect score: 2100
Sequence: 1 caatattcaagctatacca.....gaaggaaaaaaaaaaaaaa 2100

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues
Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*

44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: em_est35:*
83: em_est36:*
84: em_est37:*
85: gb_est48:*
86: gb_est49:*
87: gb_est50:*
88: gb_est51:*
89: gb_est52:*
90: gb_est53:*
91: gb_est54:*
92: gb_est55:*
93: gb_est56:*
94: gb_est57:*
95: gb_est58:*
96: gb_est59:*
97: em_est60:*
98: em_est61:*
99: em_est62:*
100: em_est63:*
101: gb_est64:*
102: gb_est65:*
103: gb_est66:*
104: gb_est67:*
105: gb_est68:*
106: em_est69:*
107: em_est70:*
108: em_est71:*
109: em_est72:*
110: em_est73:*
111: em_est74:*
112: em_est75:*
113: gb_est76:*
114: gb_est77:*
115: em_est78:*
116: gb_est79:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	516	24.6	547	73	AW540915	AW540915 C0138F11-
2	501.4	23.9	542	38	AI324866	AI324866 m181d02.x
3	499.6	23.8	507	38	AI324242	AI324242 m181d02.y
4	497.2	23.7	502	20	AA038079	AA038079 m181d02.r
5	385.4	18.4	483	79	AW659798	AW659798 97502 MAR
6	359.4	17.1	472	32	AA893267	AA893267 EST197070
7	354.8	15.9	641	46	AI952795	AI952795 w550709.x
8	319.4	15.2	420	32	AA900024	AA900024 UT-R-E0-d
9	274.4	13.1	361	33	AA980931	AA980931 u847605.r
10	259	12.3	396	72	AW407004	AW407004 UT-HF-BLO
11	243.8	11.6	572	79	AW657869	AW657869 93496 MAR
12	237.8	11.3	366	79	AW659454	AW659454 96640 MAR
13	229.6	10.9	543	79	AW654760	AW654760 104853 MAR
14	229.6	10.9	553	79	AW654759	AW654759 104851 MA
15	220.4	10.5	488	42	AI633481	AI633481 lh62f04.x
16	216.8	10.3	429	91	W85898	W85898 zh56g07.s1
17	193.6	9.2	273	81	C84657	C84657 C84657 oste
18	187.2	8.9	219	60	AV312229	AV312229 AV312229
19	144.6	6.9	172	59	AV251387	AV251387 AV251387
20	134	6.4	340	46	AI964388	AI964388 EST269502
21	133.8	6.4	420	32	AA866798	AA866798 vx90c03.r
22	127.8	6.1	163	87	R00032	R00032 ye70b04.s1
23	124	5.9	312	72	AW451487	AW451487 UT-H-BI3-
24	120.8	5.8	344	91	W85984	W85984 zh56g07.r1
25	120.6	5.7	209	37	AI234265	AI234265 EST230953
26	113	5.4	184	72	AW446306	AW446306 84924 MAR
27	111.2	5.3	368	46	AI926042	AI926042 w44c10.x
28	102.6	4.9	489	27	AA541844	AA541844 V101h01.r
29	100.6	4.9	496	42	AI615392	AI615392 V101h01.y
30	100.2	4.8	370	45	AI877165	AI877165 uc55g07.r
31	95.2	4.5	412	21	AA082133	AA082133 z042a08.r
32	87.8	4.2	279	45	AI865169	AI865169 wK09a09.x
33	82.8	3.9	316	63	AW071376	AW071376 x865e12.x
34	77.8	3.7	258	89	T47780	T47780 yb17a06.s1
35	73.4	3.5	1339	46	AI964216	AI964216 EST269330
36	68.8	3.3	297	71	AW357255	AW357255 40198 MAR
37	68.4	3.3	341	87	R00125	R00125 ye70b04.r1
38	65.6	3.1	495	39	AI450014	AI450014 mr87e10.x
39	63.4	3.0	999	46	AI964330	AI964330 EST269444
40	62.8	3.0	672	46	AI964387	AI964387 EST269501
41	61.4	2.9	824	38	AI374583	AI374583 MEST2-H1.
42	58.2	2.8	997	122	CNS005TE	AI606767 Drosophila
43	56.8	2.7	597	70	AW276523	AW276523 xrl5a05.x
44	55.2	2.6	1379	46	AI964277	AI964277 EST269391
45	52.4	2.5	502	62	AV395569	AV395569 AV395569

ALIGNMENTS

RESULT 1
LOCUS AW540915/c
DEFINITION C0138F11-3 Mouse E7.5 Extraembryonic Portion CDNA Library Mus

ACCESSION AW540915
VERSION AW540915.1 GI:7183332
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 547)
AUTHORS Kargul,G.J., Wang,X., Plao,Y., Dudekula,D., Lim,M.K., Tanaka,T., Sano,Y., Pantano,S., Grabovac,M.J. and Ko,M.S.H.
TITLE Systematic Analyses of NTA Mouse E7.5 Extraembryonic Portion CDNA Library
JOURNAL Unpublished (2000)
COMMENT On Jan 6, 2000 this sequence version replaced gi:6677269.
CONTACT: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdkn@igsun.grc.nia.nih.gov
Plate: C0138 Row: F Column: 11
Seq primer: -21M13 Forward
High quality sequence stop: 547
POLY-A-Yes.

FEATURES
source
1..547
/location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C0138F11"
/clone_1lb="Mouse E7.5 Extraembryonic Portion CDNA Library"
/sex="unknown"
/dev_stage="7.5dpc Embryo"
/lab_host="DH10B"
/note="Vector: pSPORT1 (Gibco/BRL Life Technology); site_1: SalI; site_2: NotI; Total RNAs were extracted from 5 EPC. The double-stranded cDNA was synthesized by Gibco's kit with an Oligo(dT) primer [NotI primer-adaptor from GibcoBRL]
[5'-GCAGTATGTCATGATCGAGCGGCGCCCTTTTCTTTT-3']
from 0.8ug of mRNA. The double-stranded cDNAs were treated with T4 DNA polymerase and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker LL-SalI (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Xiaohong Wang."

BASE COUNT 114 a 156 c 119 t
ORIGIN

Query Match 24.6%; Score 516; DB 73; Length 547;
Best Local Similarity 98.7%; Pred. No. 3.2e-124;
Matches 541; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

QY 1539 ctactatgacagaggtgacccacatgattggcaagccctagatcacaagccctctcgcg 1598
|||||
Db 547 CTACTATGACAGAGGA-GTGACCCACACTGATTTGGCAGCCCTCAGCCCTCTGTGG 489
|||||

QY 1599 tctgataaagaggtctctctgagctgtacatggaagtcacaaacacacactctctcc 1658
|||||
Db 488 TGTGATTAAGAGGTTCTCTGGGCTGCTACATGGAAGTCCCAAGACACACTTCTGCTCC 429
|||||

QY 1659 tctgtcttcaagagactctgactccaccctgagcggaatgagttgtctcaagcgtc 1718
|||||

Db	428	TGCTGGTTCCACAGACACTGTACTGCCACCCTTGAGCGGAATGAGTGGT-TACGCATC	370
OY	1719	cattgaagtcgacggcgcaaccaggaaaccttaactcatcagcccaaggataccoggcact	1778
Db	369	CATCGAAGTCGAGGCGCACCCAGGAAACCTTAACCTCATACGCCAGCATACCGGGCACT	310
OY	1779	ctaacatactatgcacagaattctgatagctcgagcaattcccgaggaaactctccgac	1838
Db	309	CTAACGACTACACCCGCACAGATTCTATGACTGACATTTTCGGGGAGAATCCTGGC	250
OY	1839	ggtcalctctygaaggggaagatatgctcgtgtgactgtctgagcggaaacgacaagtgtctt	1898
Db	249	GGTCAATCTCGGAAGGGAGGATGCTGCTGGAGACTTTGGACGGACAACGACTGGCTT	190
OY	1899	tgtccctgggtcgtacttgyagaagctctgaggaagagctagcaagctcacatactccc	1958
Db	189	TGTCCCTGGGTGCTACTTGTGAGAACCTCGAGGAAGGCTACAGATCTCCACATATCTCC	130
OY	1959	gccctgacatgtagtccaagacgtttcttccatataccgccccagagcttacaggggcag	2018
Db	129	GCCCTGACTGTAGGTCAGAGACTGTTCTTTCCATPCACCCGACGGCTTACGGGGCAG	70
OY	2019	aaccagaccggctgtgtcgtcggcatgtgctggtgctgtgctgctcctaataatgtccc	2078
Db	69	AACCAAGCCGGGTGGTGGTGGGAATGGCTGGGTGCTGCTACTCTCAATAATGTCTCC	10
OY	2079	cagaagga 2086 	
Db	9	CAGAAGGA 2	
RESULT 2			
LOCUS	A1324866/C		
DEFINITION	A1324866 542 bp mRNA EST 23-DEC-1998		
VERSION	m1d102.xl Soares mouse p3NNF19.5 Mus musculus cDNA clone		
KEYWORDS	IMAGE:472995.3; similar to TR:P97814 P97814 PEST PHOSPHATASE INTERACTING PROTEIN.; ; mRNA sequence.		
ACCESSION	A1324866		
VERSION	A1324866.1 GI:4059295		
SOURCE	EST.		
ORGANISM	house mouse. Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 542) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le'M., Martin,J., Morris,B., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
TITLE	The WashU-HHMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	On Jun 15, 1998 this sequence version replaced gi:3227103. Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicinep 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royally-free through LMU; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGJ:283739 This clone was previously sequenced on the 5' end only, this new data is from the 3' end Possible reversed clone: similarity on wrong strand High quality sequence stop: 473. Location/Qualifiers 1..542 /organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:472995" /clone_1lb="Soares mouse p3NNF19.5"		
FEATURES			
SOURCE			

	/dev/stage="19.5 dpc total fetus"
	//lab_host="DH10B (ampicillin resistant)"
	/note="Vector: pRT3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGAGGCGGGCCGCATTGTCTTTTTTTTTTTT 3'], Tm=68°C.
	The Not I and Eco RI sites of a modified pRT3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Felima Ronaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)." /
BASE COUNT	112 a 152 c 155 g 123 t
ORIGIN	
Query Match	23.9%; Score 501.4; DB:38; Length 542;
Best Local Similarity	97.0%; Pred.No.2.le-120;
Matches 523:	Conservative 0; Mismatches 11; Indels 5; Gaps 1;
Oy	attggcagccctcatgacatcacgccttcgtcggttgatgaagaaggttctctgtagtcta 1626
Dd	ATTGGCAGGCCCTTAGGGTCACAGCCCTCCTGTGATTAAGAAGTTCTCTGGCTGCTA 483
Oy	cattgaaagtcccccaaacacacacctctctgctctgctcttcacagaaacttgatccc 1686
Dd	CATGGAAGTCCCAAACACACACACTTCTGCTCCTGCTGCTTCACAGAGACTCTGACTCCC 423
Oy	acccttgagcgggaatgatgttgtcttaacgatalcatacgaaygcagcgcaaccaggaaac 1746
Dd	ACCCTTGAGCGGAATGAACTGTGCTTACGCAATCATCGAAGTGCACAGCACCCAGGGAAC 363
Oy	cttaactaatcaagccaggaataccgggccaactctcagaactacaactcagaagaattctgat 1806
Dd	CTTAACATCATCAGCCCAAGSACTACCGGGCAGCTTACGACTACACCGCACAGAAITTTGAT 303
Oy	gagcttggaatttccgcyggagacalcccttygcycgtcalcccttgaaaagggagatgcttg 1866
Dd	GAGCTGGAATTTCCGGCGGAGACAATCCTGGCGGTGCATCCTTGGAAGGGAGATGGCTGG 243
Oy	tggacttgyggagcggaaacggaacgtygcttctgtcccttgagtgtaacttggagaagctc 1926
Dd	TGGACTTGAGAACGGGAMCGGACMACGTGGCTTGTTCCTTGGGTCTCTACTTGGAGAAAGCTC 183
Oy	tgaagaaagsgctcagagctccacatactccgcccttacttgaagtgcagagactgattc 1986
Dd	TGAGAAAAGGCTAGCAGTCTCCACATACTCCGCCCTAATCTGTGAGTCAAGAGACTGTTTC 123
Oy	ttccatcaacgcccacagagcctlaecggggc-----cagaaccaagcccgatggtcttgggc 2041
Dd	TTTTCATCATCAGCCGCCAGGCGCTCAGCGGGCGTTATCAGAAACCAAGCCCGTGGTGGGA 63
Oy	atgggcctggtgctgctgctactctcaaataatgctctccagaagaagaaaaaaaaaaaaa 2100
Dd	ATGGGCTGGGTGCTGTGCTACTCTCAATAAATGTCTCCAGAAAGGATTAACCGCGCAAAA 4
RESULT 3	
A1322422	LOCUS A1322422 507 bp mRNA EST 23-DEC-1998
DEFINITION	m.81d02.y1 Soares mouse pJNMF19.5 Mus musculus cDNA clone IMAGE:472995 y1 similar to FR:p97814 p97814 PST PHOSPHATASE INTERACTING PROTEIN.; mRNA sequence.
ACCESSION	A1322422
VERSION	A1322422.1 GI:4056851
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota: Metazoa: Chordata: Cranata: Vertebrata: Euteleostomi; Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 507)
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubouque,T.,

	Matches	499,	Conservative	0;	Mismatches	3;	Indels	0;	Gaps	0;
QY	1420	gagctctataggaagtgcgctgcgaaccttgaggctgtatgtvgaagtgtaacaaat	1479							
Db	1	GAGCTCTATGAGGAAGTGC	CGCGCTGACCCCTTGAGGCGCTTGATGTGTGAAGGTGACATCAAT	60						
QY	1480	ggcttcacacagctccaagaagaactctgcaagagagccccagctccggttcctatragaac	1539							
Db	61	GGCTTCATCCGATCCAAAGACATGGCAGAGAGCCCCAGAGCTCCGCTCCCTTATCAGAAC	120							
QY	1540	tactatgcagggaaggtgtaccccaactgtattgcaagccctaaagacacccctcctgcggt	1599							
Db	121	TACTATGACAGGAGAGTGA	CCCACTGATGTGGACACCTTAGAGTCCAGCCCTCCGTGTGT	180						
QY	1600	gtgataaagaggttcctctgtagctgtcatatgtgaagtlcccaagaacacacactctgcct	1659							
Db	181	GTGATTAAGAGAGTTCCTGCGCTCTACATATGAAATGCCAAGACCACACCTTCCTCCT	240							
QY	1660	gctgcttccacaagagactctgaactcccaacctctgagcggaatgaattggtctagcatcc	1719							
Db	241	GCTGCTTCCACAGAGAACCTTGACTCCACCCCTGAGCGGAATGAGATTGTCATACCACATCC	300							
QY	1720	atcgagagtcgacgaccccaagagaaccttaactataagccccagagataccgggcatc	1779							
Db	301	ATCGAGATGCAAGCGCACCAAGGAGAACCTTATCTATATGACCCAGAGACTACCGGGCATC	360							
QY	1780	tacgactacactctgcacagaattctgtatgatgctgtgcacattccgcgggaagacalcctg	1839							
Db	361	TACGACTACACCGCACACAAATTTCTGATGAGCTGGACATTTTCCGGCGGAGACATCCTGGCG	420							
QY	1840	gtcatcctcgaaaggggaagatlgctgtgtgagactgtlgagcggaagcgaaacgtgtgctt	1899							
Db	421	GTCATCCTCGGAAGGAGAGAGATGCGTGTGAGACTGTGAGCGGAACGGACAACGTGGCTTT	480							
QY	1900	gtccctggtgctgtacttgaga	1921							
Db	481	GTCCCTGGGTGTACTTGAGA	502							

RESULT	5
LOCUS	AM659798
DEFINITION	AM659798 483 bp mRNA EST
ACCESSION	U01080
VERSION	1
KEYWORDS	Bos taurus.
SOURCE	Bos taurus.
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidæ; Bovinae; Bos.
REFERENCE	Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahrenkrug,S.C., Freiling,B.A., Rohrer,G.A. and Keefe,J.W.
AUTHORS	Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle
JOURNAL	Unpublished (2000)
COMMENT	On Jan 6, 2000 this sequence version replaced g1:6675666.

Email: smith@emil.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904-e. Vector identified by cross_match with the -mnscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTTCCGAGTCACGACG
Plate: 71 row: 1 column: 11
seq primer: ATTTAGTGACACTACTAAG.

FEATURES	SOURCE	LOCATION/QUALIFIERS
	1	483
		/organism="Bos taurus"
		/db_xref="taxon:9913"
		/clone_lib="MARC 1B0V"
		/tissue_type="pooled"
		/lab_host="DH10B"
		/note="Vector: pCMV SPOR6; Site_1: xbaI; Site_2: xhoI Library made from pooled tissue from lymph node, ovary fat, hypothalamus, and pituitary."
BASE COUNT	127 a	126 c 163 g 67 t
ORIGIN		

Query Match	18.4%	Score 385.4	DB 79	Length 483
Best Local Similarity	87.4%	Pred. No. 4.2e-90		
Matches 422	Conservative 0	Mismatches 61	Indels 0	Gaps 0
QY 943	gacagtcacacatccacagctggccctggccctgcgtctgagagagctcgagagccttgagagag	1002		
Db 1	ggcagactccacatccacactggccttgccctggccctggccgacgagcgtgggagccttgagagag	60		
QY 1003	ttccgagagagacagaaagacgcggaagaagtatgagccatcatgagccgtgtccag	1062		
Db 61	ttcccttgagagggcagaaagcagcagagaaagtaacgagcctgtcatgagccgtgtccag	120		
QY 1063	aagagcaagctgtgcctcttaagaagagccatgtggtctcaagaagcctatgaccagag	1122		
Db 121	aagaccaaagctgtccctctccacaaagagccatgagctccaaagaaacctaattgacgaag	180		
QY 1123	tgcaagagaatgcagatgatctgcagcagagccttcagacgctgtgagtgccaatgagccaccag	1182		
Db 181	tgccggagatggcgagatgacgcgtgaaacagccctttgagcccatattgacaccacagcgtcaacag	240		
QY 1183	aagcaagtagaanaagagccagaacaagccaagcctgtcaagagagtcagccacagagca	1242		
Db 241	aagcagctggagaaagcagcagcagcaaaacccaagcagctcaagagacttcagcatgagagca	300		
QY 1243	gaaagagtgtaacagccaatatatcgaaacacttgagaaagagagagagacccagatggagacag	1302		
Db 301	gagccacacatcagagcaaaacatttgacgacgtggagagaggtccggatgtgagctggagacag	360		
QY 1303	gagcaccgagactaccctgtgagagccttcacagltgcagagagatltgaccagcaccacctc	1362		
Db 361	gagcaccgccgcacacctgggagagccctttccagttgcacgagatttgaccggctgaccatctc	420		
QY 1363	cgcgaatccctgtgtggtgtgcaactgttaaccagcctccatgtcagagtgctgttcaagatgtatgag	1422		
Db 421	cgcgaactccctgtgtggtgtgcaactgtcagcaaacacagctcttcacatgacagtgctgtcaagacgatgag	480		
QY 1423	ctc 1425			
Db 481	ctc 483			

RESULT	6
AA893267/c	
LOCUS	472 bp mRNA EST 16-JUN-1998
DEFINITION	EST197070 Normalized rat kidney, Bonto Soares Rattus sp. cDNA clone RKBSE29 3' end, mRNA sequence.
ACCESSION	AA893267
VERSION	AA893267.1 GI:3020146
KEYWORDS	EST.
SOURCE	Rattus sp.
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. 1 (bases 1 to 472)
AUTHORS	Lee,N.H., Glodok,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlagage,A.R. and Adams,M.D.
TITLE	Rat Genome Project: Generation of a Rat EST (RESt) Catalog & Rat Gene Index

OY	1943	gctccacatcactccgcgttcagctgtagtgcagaacctgtttcttcataaccgccca	2002
Db	145	ggccccttggaacct--gccctccacagtgtggaggcacacagtgcgccccaacagtctc--ccca	90
OY	2003	ggcctcacggggcgccagaaccaagc	2026
Db	89	ccttcctfagggccccagaaaccaagc	66
RESULT	8		
LOCUS	AA900024/c		
DEFINITION	AA900024	420 bp mRNA	EST
	UI-R-E0-dh-d-01-0-VI.s1	UI-R-E0 Rattus norvegicus cDNA clone	04-JUL-1999
	UI-R-E0-dh-d-01-0-vi 3,	similar to gi11857711 gb U87814 MM087814	
	Mus musculus PEST phosphatase interacting protein mRNA, complete		
	cds, mRNA sequence.		
ACCESSION	AA900024		
VERSION	AA900024.1	GI:4232513	
KEYWORDS	EST.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Rattus.		
REFERENCE	I (bases 1 to 420)		
AUTHORS	Bonaldi,M.F., Lennon,G. and Soares,M.B.		
TITLE	Normalization and subtraction: two approaches to facilitate gene		
JOURNAL	discovery		
MEDLINE	Genome Res. 6 (9), 791-806 (1996)		
COMMENT	97044477		
	On Apr 7, 1998 this sequence version replaced gi:3035378.		

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mssoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult 8-day-Embryo library. cDNA library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1770653

Seq primer: M13 Forward
POLYA=No.

```

FEATURES
source
location/Qualifiers
1..420
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-dh-d-01-0-01"
/clone_1lb="UI-R-E0"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pUT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: NotI; Site_2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dt track which allows
identification of the library of origin of a clone within
the mixture."
BASE COUNT
84 a 120 c 107 g 106 t 3 others
ORIGIN

```

Query Match	15.2%	Score 319.4	DB 32	Length 420
Best Local Similarity	86.3%	Pred. No. 7	1e-73	
Matches 364	Conservative 0	Mismatches 54	Indels 4	Gaps 1
Oy	1679	tgactccaaccctgagcgaatgagtgtgtctacgcatacatcgaagtgcagcagacc	1738	

Db	420	TGGCTCCACCTTTGAGCAGAAAGAAATGGTNNATGCAATCATTTGAAGTCAGGCAACNC	361
QY	1739	agggaacctaaatcatcatcaagcccaaggataccgggactctacgactaacactcacaga	1798
Db	360	AGGGAGACCTTAACTCAGACAGCCCAAGACTACCGGGGGCTTATGACTACAGGACACA	301
QY	1799	attctgatagactgtgacatctccggcgggagacatcctgtgcggtcatcctgtgaaggga	1858
Db	300	ATTCTGATGAGCGGACATTTCTGCGGGAACATCCTGGCAGTCACTCTGMAAGGAGG	241
QY	1859	atggctggtggagactgtggaagcggaagaaagtgacttlytccctgggtcgacttgg	1918
Db	240	ATGGCTGGTGGACTGTGGAACGAACGGAAACCTGGCTTGTGCTTGCGTACTTGG	181
QY	1919	agaagctcgaagaagagctagcagctccacatacctccggccctgactgtgagtcaga	1978
Db	180	AAAGACTCTGAGAGAAAGACACAGCGGTCCACATACCTCTGCCCCGGCTGAGGTCAAGG	121
QY	1979	actgttctctcatcaacccgcccagagcgtctcaagggcgagaaaccaagcccggtgtgtc	2038
Db	120	ACTG----TTCCATATCACGCCCCGGCGCTGTGGGGGTCAACAACCATGGCTGTGCCGTG	65
QY	2039	ggcatggctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2098
Db	64	GGCATGGGATGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	5
QY	2099	aa 2100	
Db	4	AA 3	

RESULT	9
AA980931	
LOCUS	
DEFINITION	AA980931 361 bp mRNA EST 27-MAY-1998
ACCESSION	U447605.r1 Soares;mammary_gland_NDMMG Mus musculus cDNA clone
VERSION	IMAGE:1344888.5' similar to TR:P97814 P97814 PEST PHOSPHATASE
KEYWORDS	INTERACTING PROTEIN. ;, mRNA sequence.
SOURCE	AA980931
ORGANISM	AA980931.1 GI:3159467
REFERENCE	EST.
AUTHORS	house mouse.
	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus
	1 (bases 1 to 361)
	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
	Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
	Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
	Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
	Waterston,R.
TITLE	The Washu-HHMT Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Apr 7, 1998 this sequence version replaced gi:3036147.
	Contact: Marra M/Mouse EST Project
	Washu-HHMT Mouse EST Project
	Washington University School of Medicinep
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
	Tel: 314 286 1800
	Fax: 314 286 1810
	Email: mouseest@watson.wustl.edu
	This clone is available royalty-free through LMD ; contact the
	IMAGE Consortium (info@image.lnl.gov) for further information.
	MG1:698680
	Seq primer: -28m13 rev2.ET from Amerisham
	High quality sequence scop: 276.
FEATURES	
Source	1..361

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="1349888"
/clone_lib="Soares_mammary_gland_NBMG"

```

```

/sex="male"
/issue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCCGCAATGTTTTTTTTTTTTTTTTTTT
3'] double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."

```

BASE COUNT
ORIGIN

87 a 90 c 108 g 76 t

Query Match 13.1%; Score 274.4; DB 33; Length 361;
Best Local Similarity 99.6%; Pred. No. 3.9e-61;
Matches 275; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1390 cagcttcacagcagctgtgtgaagatgaagctctatgaagagtgagctgacccct 1449
|||||
DB 1 CAGCTCTCCATGACAGTGTGTCAGAGATGATGAGCTCTATGAGAAAGTCCGCTGACCTT 60

QY 1450 gaggctgtgtgtgtgaagtgatcatcatgtctcatccagccaagagctgacaga 1509
|||||
DB 61 GAGGCTGTGTGATGTGGAAGTGTGATCATCATGCTTCATCCAGTCCAGACAGCAGTGGCAGA 120

QY 1510 gagcccccagctccggtgcttcatcagaactatgacagagagtgagcccaatgatt 1569
|||||
DB 121 GAGCCCTCCAGCTCCGGTCCCTTATCAGAACTACTATGACAGGAGTGCACCCCACTGATT 180

QY 1570 ggcgcgcctagctacccagcctctctggtgtgtgtataaagagttctctgggtcttacc 1629
|||||
DB 181 GGCAGCCCTACATCCACCTCTCGCGTGTGATAAAGAGTTCTCTGGGCTCTACAT 240

QY 1630 ggaagtcacagacacacccctctgctctgctgct 1665
|||||
DB 241 GGAAGTCCCAAGACACACCTTCTGCTCTCTGCT 276

RESULT 10

AM407004 396 bp mRNA EST 16-FEB-2000

LOCUS UT-HF-BLO-adf-g-08-0-UI.r1 NIH_MGC_37 Homo sapiens cDNA clone
DEFINITION IMAGE:3061310 5', mRNA sequence.

ACCESSION AM407004.1 GI:6926061

VERSION AM407004.1

KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 396)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT On May 20, 1999 this sequence version replaced gi:4880298.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 forward.

FEATURES
source

Location/Qualifiers
1.396

BASE COUNT
ORIGIN

95 a 130 c 111 g 60 t

Query Match 12.3%; Score 259; DB 72; Length 396;
Best Local Similarity 80.3%; Pred. No. 4.2e-57;
Matches 317; Conservative 0; Mismatches 75; Indels 3; Gaps 1;

QY 1415 atgatgaccttatgaggaagtgctgctgaccccttgagggctgtgatgtgaagtgaca 1474
|||||
DB 2 ACGAGGAGCTCTACGAGAAAGTGGGCTGACGCTGAGAGGCTGACGATGAGCGGACA 61

QY 1475 tcaatgcttcatcagctcagaagacactgacagagagcccaagctcgtgcttacc 1534
|||||
DB 62 TCGACATTTTCATCCAGGCCAAGACAGCGGACAGAGAGCCCGCTCGGCTCCCTACC 121

QY 1535 agaactactatgacagagagtgagcccaatgattggcagccctagatcacaagcctct 1594
|||||
DB 122 AGACTATTTACGATCGGAGAGTACCCCGCTGACACAGACACCCCTGGCATACAGCGGCT 181

QY 1595 gcggtgtgataaagaggttctctggtgtgtgtatgagagttcccaagacacctt--- 1651
|||||
DB 182 GTGCGAATGATTAAGATGTTCTCTGAGCTGTGACAGGAAATCCCMACACATCTGTTG 241

QY 1652 ctgctcctgctgtctcacaagagactctgactccaccccttgagagagtgattgct 1711
|||||
DB 242 CAGCTTGTGCTGCTCCACAGAGACCTGACCCCGACCGGAGGATGAGGCTGTCT 301

QY 1712 acgatccatcagaagtgacagcgacccaggaaccttaactcatcagccagactacc 1771
|||||
DB 302 ACACAGCCATCGCAGTGCAGAGATACAGGAAACCCCGCTCACCGCAGAGATACC 361

QY 1772 ggcgcctcctagctacacgtcagacagattctgatt 1806
|||||
DB 362 GGGCGCTCTACGATTATACAGCGAGAACCCAGAT 396

RESULT 11

AM657869 572 bp mRNA EST 05-APR-2000

LOCUS 93496 MRC 1BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION AM657869.1 GI:7423695

VERSION AM657869.1

KEYWORDS EST.

SOURCE Bos taurus.
ORGANISM Bos taurus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 572)
AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
Keeler,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smithemil.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.

PCR primers
FORWARD: AGCAACAGCTATGACCAT
BACKWARD: GTTTCACAGTCACGAGC
Plate: 79 row: C column: 7
Seq primer: ATTAGGTGACACTATAG.

FEATURES

source
1. .572
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 139 a 142 c 186 g 105 t
ORIGIN

Query Match 11.6%; Score 243.8; DB 79; Length 572;
Best Local Similarity 87.8%; Pred. No. 4.3e-53;
Matches 266; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 895 tcccttgagacctccttctgactccctgaagcagcaagaagtgtggcagctgcacac 954
DB 270 TCCCTGAGGCGCTCCTTGTACTCTCTGAAGCAGCAAAATGGAATGTGGCGCTCCAC 329
QY 955 atcagcgtgccccttgccttgctgaagagcgtgcgagccttgagagatccgagaga 1014
DB 330 ATCCAGCTGGCCCTGCGCTGCGAGAGCTGCGAGCTGGAAGAGTTCGTGCGAGAG 389
QY 1015 cagaagaagcagcagaagaatgatagcaccatcagcagctgtccagaagaagaagt 1074
DB 390 CAGAAGAGCAGAGAGAGAGAGTATGAGAGCTGTATGAGAGAGTCCAGAGAGCAAGCTG 449
QY 1075 tccgtctcaagaagaagcagcagcagcagcagcagcagcagcagcagcagcagc 1134
DB 450 TCGCTCCACAGAGAGCGCATGAGTCCAAAGAGAGCTATGAGAGAGAGTCCGCGAGTGC 509
QY 1135 gatgatgttgagcagcagccttgagcgtgtgagtgccaatgagccaccagaagaagtaga 1194
DB 510 GATGAGCTGGAACAGGCGCTTGTAGCGATTAGCACCACAGTCAACAGAGAGAGTGGAG 569
QY 1195 aag 1197
DB 570 AAG 572

RESULT 12
AM659454 366 bp mRNA EST 05-APR-2000
LOCUS 96640 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AM659454
VERSION AM659454.1 GI:7425281
KEYWORDS EST.
SOURCE Bos taurus.
ORGANISM Bos taurus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 366)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Fekling,B.A., Rohrer,G.A. and
Keefe,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle

JOURNAL Unpublished (2000)
COMMENT On Jan 6, 2000 this sequence version replaced gi:1675298.
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smithemil.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.

PCR primers
FORWARD: AGCAACAGCTATGACCAT
BACKWARD: GTTTCACAGTCACGAGC
Plate: 74 row: A column: 5
Seq primer: ATTAGGTGACACTATAG.

FEATURES

source
1. .366
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 76 a 109 c 123 g 58 t
ORIGIN

Query Match 11.3%; Score 237.8; DB 79; Length 366;
Best Local Similarity 83.8%; Pred. No. 1.4e-51;
Matches 269; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 1241 cagaagaagtgtacagcagaatactgaaacaactgagagagagcagcagatggagagc 1300
DB 46 CAGAGCGAATACATAGGCAAGCAATTTAGAGAGCTGGAAGAGGTGAGAGAGAGC 105
QY 1301 aggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1360
DB 106 AGGAGCAGCCCGCCAGCTCGGAGGCGCTTCAGATTGACAGCGCTGACCAATCC 165
QY 1361 tccgcaatgccccttgggtgagcagcagcagcagcagcagcagcagcagcagcagc 1420
DB 166 TCCGCAACTCCGTGTGGGTGACATGCAACAGCTCTCCATGCAATGGCTCAAGAGCATG 225
QY 1421 agctctatgagagagtgagcagcagcagcagcagcagcagcagcagcagcagcagc 1480
DB 226 AGCTCTAGCAAGAGGTGCGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGG 285
QY 1481 gctcatcagctcagaagaagcagcagcagcagcagcagcagcagcagcagcagcagc 1540
DB 286 GCTTCATCCAGGCCAAGAGAGCAGCTGAGCCCGCAGCCGCGCTTACCAAGAACT 345
QY 1541 actatgacagagagagagcc 1561
DB 346 ACTATGATCGGAGGTGCGCC 366

RESULT 13
AM654760 543 bp mRNA EST 05-APR-2000
LOCUS 104853 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AM654760
VERSION AM654760.1 GI:7420586
KEYWORDS EST.
SOURCE Bos taurus.
ORGANISM Bos taurus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 543)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Fekling,B.A., Rohrer,G.A. and
Keefe,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle

TITLE
Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle
JOURNAL
On Oct 8, 1998 this sequence version replaced gi:3729921.
COMMENT
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross_match with the -mnscore 18 and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 100 row: D column: 4
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
1. .543
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."
BASE COUNT
103 a 188 c 154 g 98 t
ORIGIN

Query Match
Best Local Similarity 10.9%; Score 229.6; DB 79; Length 543;
Matches 290; Conservative 0; Mismatches 79; Indels 3; Gaps 1;

1609 aggtctctggtgctgctacatggaagtcacagacacacacctct--gtctcgtgctgct 1665
|||||
105 AGGTTTCTGGGCTGCTGATGAGTCCAGACACACACTGTGACGACGCTTTCGACGCC 164
|||||
1666 tcacagagactctgactccacccctgagcgaatgagtgtctacgcatccatcgaa 1725
|||||
165 TCCACAGAGACCCCTGACTCTTACCCCTGACCGGATGAGGGTGTCTACGCTGCATCGCC 224
|||||
1726 gtgcaggcgaccggaggaacctaactcaagcccaagcgtacgcgggacatctagac 1785
|||||
225 GTGCAGAGAGGCGCAGGACCCGCCACCATGCGAGGCCAGCTATAGGGCAGCTATGAC 284
|||||
1786 taacactgacagaattctgtagtgcgtgacatcttcgcggaagacatccctgcggtc 1845
|||||
285 TACACAGCCCGAGACTCTGAGAGCTGAACATCTCTCGGGGAGACATCTGAGAGTCA 344
|||||
1846 ctggaagaggagatgctgctgtagtgcgtgagcggagaaacagctgctgtccct 1905
|||||
345 CTGGAAGAGGAGGAGCGCTGCTGAGACGCTACAAAGCGGAGCGCTGCTTGTGCTCT 404
|||||
1906 gggtcgacttggagaaagcctgtagcagctccacatccacatcccgccctga 1965
|||||
405 GGCTCCTACTCTGGAGAAATTTAAGAGAGACAGTAGCCCTCTCTGAGACTTCTCTGC 464
|||||
1966 ctgtgaggtcag 1977
|||||
465 CAGTGGGGCCAG 476
|||||

RESULT 14
AM654759 553 bp mRNA EST 05-APR-2000
LOCUS
DEFINITION 104851 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AM654759
VERSION AM654759.1 GI:7420585

KEYWORDS EST.
SOURCE Bos taurus.
ORGANISM Bos taurus
REFERENCE
AUTHORS
1 (bases 1 to 553)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and Keeler,J.W.
TITLE
Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle
JOURNAL
On May 18, 1998 this sequence version replaced gi:3137243.
COMMENT
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross_match with the -mnscore 18 and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 100 row: D column: 2
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
1. .553
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."
BASE COUNT
105 a 190 c 156 g 102 t
ORIGIN

Query Match
Best Local Similarity 10.9%; Score 229.6; DB 79; Length 553;
Matches 290; Conservative 0; Mismatches 79; Indels 3; Gaps 1;

1609 aggtctctggtgctgctacatggaagtcacagacacacacctct--gtctcgtgctgct 1665
|||||
105 AGGTTTCTGGGCTGCTGATGAGTCCAGACACACACTGTGACGACGCTTTCGACGCC 164
|||||
1666 tcacagagactctgactccacccctgagcgaatgagtgtgtctacgcatccatcgaa 1725
|||||
165 TCCACAGAGACCCCTGACTCTTACCCCTGACCGGATGAGGCTGTCTGCGCATCGCC 224
|||||
1726 gtgcaggcgaccggaggaacctaactcaagcccaagcgtacccgggacatctagac 1785
|||||
225 GTGCAGAGAGGCGCAGGACCCGCCACCATGCGAGGCCAGGAGCTATAGGGCAGCTATGAC 284
|||||
1786 taacactgacagaattctgtagtgcgtgacatcttcgcggaagacatccctgcggtc 1845
|||||
285 TACACAGCCCGAGACTCTGAGAGCTGAACATCTCTCGGGGAGACATCTGAGAGTCA 344
|||||
1846 ctggaagaggagatgctgctgtagtgcgtgagcggagaaacagctgctgtccct 1905
|||||
345 CTGGAAGAGGAGGAGCGCTGCTGAGACGCTACAAAGCGGAGCGCTGCTTGTGCTCT 404
|||||
1906 gggtcgacttggagaaagcctgtagcagctccacatccacatcccgccctga 1965
|||||
405 GGCTCCTACTCTGGAGAAATTTAAGAGAGACAGTAGCCCTCTCTGAGACTTCTCTGC 464
|||||
1966 ctgtgaggtcag 1977
|||||
465 CAGTGGGGCCAG 476
|||||

THIS PAGE BLANK (USPTO)

OM of: US-09-068-377-1 to: EST:* out_format : pfs

Date: Sep 15, 2000 7:42 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+ .p2n.model -DEV=xlp
-Q=/cgn2.1/USPFO.spool/US09068377/unal.14092000.165620.22394/app_query.fasta.1.477
-DB=EST -OPT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-DEEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-LIST=45 -LOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0
-ALIGN=15 -MODE=LOCAL -OUT=pfm -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09068377.ecgnl.1.1576 -NCPU=6
-ICPU=3 -LONGLOGS -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-068-377-1

Query length: 415

Database: EST:*

Database sequences: 5247842
Database length: -2090053206
Search time (sec): 647.830000

Score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_est1:AA038079	876.00	1579.91	3.6e-79	502	AA038079 mi81d02.r1 Soares mous
gb_est19:AI322422	871.00	1570.71	1.2e-78	507	AI322422 mi81d02.r1 Soares mous
gb_est45:AM659798	764.00	1376.65	7.5e-68	483	AM659798 97502 MARC IBOV Bos t
gb_est27:AI952795	637.50	1143.63	7.2e-55	641	AI952795 w50709.x1 NCI_CGAP_Ut
gb_est1:AI322486	621.00	1115.40	2.7e-53	542	AI322486 mi81d02.r1 Soares mous
gb_est43:AM407004	614.00	1102.52	1.4e-52	547	AM407004 C0138F11-3 Mouse E7.5
gb_est42:AM659454	531.50	955.97	2.0e-44	396	AM659454 UI-HF-BH0-adj-d-08-0-0
gb_est13:AM693267	516.00	928.66	6.8e-43	366	AM693267 E9640 MARC IBOV Bos t
gb_est14:AA80931	501.50	899.56	2.8e-41	472	AA80931 E9640 MARC IBOV Bos t
gb_est4:AM657869	488.50	878.76	4.1e-40	361	AM657869 93456 MARC IBOV Bos t
gb_est45:AM654760	456.00	814.76	1.5e-36	572	AM654760 104853 MARC IBOV Bos t
gb_est45:AM654760	446.50	798.06	1.3e-35	543	AM654760 104853 MARC IBOV Bos t
gb_est13:AA000024	400.00	716.22	4.6e-31	420	AA000024 UI-R-E0-dh-d-01-0-0-UI-S
gb_est54:W85898	393.00	703.27	2.4e-30	429	W85898 zh56907.r1 Soares-fetal
gb_est23:AI633481	378.50	675.54	8.5e-29	488	AI633481 E82F04.x1 NCI_CGAP_Ut
gb_est27:AI164388	372.00	662.65	4.4e-28	540	AI164388 E82F04.x1 NCI_CGAP_Ut
gb_est15:W85984	323.50	579.23	2.0e-23	344	W85984 zh56907.r1 Soares-fetal
gb_est47:CA8657	315.00	566.22	1.0e-22	273	CA8657 C8657 osteoclast subtra
gb_est8:AA51844	315.00	560.05	2.3e-22	489	AA51844 vj01b01.r1 Barstead mc
gb_est23:AI615192	301.50	537.31	4.2e-21	412	AI615192 vj01b01.r1 Barstead mc
gb_est50:R00032	265.00	480.75	6.0e-18	163	AA082133 zn42a08.r1 Stratagene
gb_est42:AA646306	242.00	437.65	1.5e-15	184	AA646306 84924 MARC IBOV Bos t
gb_est27:AI926042	227.00	401.64	3.0e-13	420	AA926042 vx90c03.r1 Soares fetal
gb_est42:AA651487	211.00	375.69	4.3e-12	312	AA651487 ui-h-B3-aii-g-03-0-0-UI
gb_est50:R00125	177.50	313.83	1.2e-08	341	R00125 ye70b04.r1 Soares fetal
gb_est41:AM659684	174.00	300.63	6.5e-08	651	AM659684 QV0-BT0041-261199-059
gb_est26:AI187165	171.00	293.15	6.0e-08	370	AI187165 uc55907.r1 Soares thym
gb_est28:AI134987	170.00	291.41	1.6e-07	646	AI134987 DKFZP762H1411.r1 762
gb_est22:AI563885	163.00	281.51	7.5e-07	599	AI563885 tn31a03.x1 NCI_CGAP_Bd
gb_est47:CA90739	162.00	282.05	7.0e-07	479	CA90739 C90739 Dictyostelium dls
gb_est21:AI513134	160.50	275.74	1.6e-06	672	AI513134 LD43318.Sprime LD Dros
gb_est36:AA782344	159.50	275.56	1.6e-06	576	AA782344 AV382344 Halocynthia
gb_est38:AA116912	154.00	267.50	4.7e-06	493	AA116912 f121d06.r1 Sugano Kaw
gb_est20:AI454404	153.50	266.14	5.4e-06	500	AI454404 UI-R-ED-cd-c-04-0-0-UI-S
gb_est45:AM632724	153.00	266.75	7.3e-06	575	AM632724 92825 MARC IBOV Bos t
gb_est42:AA421808	153.00	261.40	6.9e-06	718	AA421808 f186d12.y1 Sugano Kaw
gb_est28:AA782752	151.00	264.39	6.7e-06	384	AA782752 4A3A-ABF-07-R Anophe
gb_est37:AA071376	149.00	262.81	8.2e-06	316	AA071376 xa65e12.x1 NCI_CGAP_HS

gb_est26:AI865169 - 147.00 260.49 1.1e-05 279 i AI865169 wK09a09.x1 NCI_CGAP
gb_est1:AA030355 + 142.50 249.50 4.5e-05 364 i AA030355 mi08h12.r1 Soares m
gb_est11:AA771236 + 142.50 245.87 7.3e-05 515 i AA771236 vt17a02.r1 Barstead
gb_est15:AI006646 + 142.50 241.96 0.0001 742 i AI006646 ue16n06.y1 Sugano m

seq_name: gb_est1:AA038079

seq_documentation_block:

LOCUS AA038079 502 bp mRNA EST 28-AUG-1996

DEFINITION mi81d02.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:472995 5' similar to SW:MTSC.ACACA P10569 MTOSTIN IC HEAVY

CHAIN: [1] ; mRNA sequence.

ACCESSION AA038079 GI:1513518

VERSION AA038079

KEYWORDS EST.

SOURCE Mus musculus.

ORGANISM Mus musculus.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 502)

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Schellenberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

```

|||||
1 GAGCTCATGAGGAAGTCGGCTGACCTTGAGGGCTGTGATGTGAAG 50
263 yasp1eansnlypheiieglnserlysserthnglyargluppropa 280
51 TGACATCAATGGCTTCATCCAGTCCAGACAGCATGGCAGAGCCCGCAG 100
280 laprovalprotyrglnasnrytyraspargluvalthproleulle 296
101 CTCGGGTCCCTTATCAGAACTACTATGACAGGAGTACCCCATGATT 150
297 glyserproserileglnprosercysglyvalilelysarpheserel 313
151 GGAGAGCCCTACGTCACGCCCTCTGCTGTGTGATAAAGAGTTCTCTGG 200
313 ylenleuhisglyserprolysthrthproseralaproalaasert 330
201 GCTCTACATGGAAGTCCCAAGACACACCTTCTGCTCTCTGCTCTCCA 250
330 hrgluthleuthrprothrpogluaragasnglualeuvaltyralaser 346
251 CAGAGACTCTGACTCCACCCCTGAGCGGAATGAGTTGCTACGACATCC 300
347 lilegluvalglnalathrglnglyasnleuanserseralaglnaspty 363
301 ATCGAAGTGCAGGCGACCCAGGAAACCTTAATCATCAGCCAGACTA 350
363 fargalaleuthyrasptyrthrallaglnasnseraspcluleuaspiles 380
351 CCGGGCACTTACGACTACACCGCACAGAAATTCGATGAGCTGGACATTT 400
380 eralaglyasp1leuualavalileleuglnglyaspolytrptp 396
401 CCGGGGAGACATCCTGGCGGTCATCTGGAAGGGAGAGTGGCTGG 450
397 thryalgluaragasnglyglnarglyphevalproglyserlyleugl 413
451 ACTGTGAGCGGAGACGACACGTGGCTTGTCTGCTGGTGTACTTGA 500
413 u 413
501 G 501

```

seq_name: gb_est19:AI322422

seq_documentation_block: 507 bp mRNA

LOCUS AI322422 mi81d02.y1 Soares mouse p3NMf19.5 Mus musculus CDNA clone

DEFINITION IMAGE:472995.5 similar to TR:P97814 P97814 PEST PHOSPHATASE

ACCESSION AI322422.1 GI:4056851

VERSION AI322422.1

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

AUTHORS Maira,M., Hillier,L., Allen,M., Bowles,M., Dietlich,N., Dubuque,T.,

Tselsing,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Maira M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LNM; contact the

IMAGE Consortium (info@image.jiml.gov) for further information.
 MGI:283739
 This read is a RESPONSE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 correct orientation)
 Putative full length read
 vector to vector length is 674
 Possible reversed clone: similarly on wrong strand
 Seq primer: -40RP from Gibco
 High quality sequence stop: 475.

FEATURES

source location/Qualifiers
 1..507
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:472995"
 /clone_1lb="Soares mouse p3NMf19.5"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Vector: pT73D (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand CDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTATCCAAATCGAAGTGGAGGCGCGCATTTTTTTTTTTT 3']
 double-stranded CDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldo. RNA was kindly provided by
 Dr. Minou Ko (Wayne State University)."
 BASE COUNT 113 a 142 c 146 g 105 t 1 others
 ORIGIN

alignment_scores:
 Quality: 871.00 Length: 169
 Ratio: 5.216 Gaps: 0
 Percent Similarity: 98.817 Percent Identity: 98.225

alignment_block:
 us-09-068-377-1 x AI322422 ..

Align seg 1/1 to: AI322422 from: 1 to: 507

```

247 gluleuthrglngluvalargleuthrleuglnglycysaspvalglugl 263
|||||
1 GAGCTCATGAGGAAGTCGGCTGACCTTGAGGGCTGTGATGTGAAG 50
263 yasp1eansnlypheiieglnserlysserthnglyargluppropa 280
51 TGACATCAATGGCTTCATCCAGTCCAGACAGCATGGCAGAGCCCGCAG 100
280 laprovalprotyrglnasnrytyraspargluvalthproleulle 296
101 CTCGGGTCCCTTATCAGAACTACTATGACAGGAGTACCCCATGATT 150
297 glyserproserileglnprosercysglyvalilelysarpheserel 313
151 GGAGAGCCCTACGTCACGCCCTCTGCTGTGTGATAAAGAGTTCTCTGG 200
313 ylenleuhisglyserprolysthrthproseralaproalaasert 330
201 GCTCTACATGGAAGTCCCAAGACACACCTTCTGCTCTCTGCTCTCCA 250
330 hrgluthleuthrprothrpogluaragasnglualeuvaltyralaser 346
251 CAGAGACTCTGACTCCACCCCTGAGCGGAATGAGTTGCTACGACATCC 300
347 lilegluvalglnalathrglnglyasnleuanserseralaglnaspty 363
301 ATCGAAGTGCAGGCGACCCAGGAAACCTTAATCATCAGCCAGACTA 350
363 fargalaleuthyrasptyrthrallaglnasnseraspcluleuaspiles 380
|||||

```

351 CCGGCGACTCTACGACTACACCGACAGAAATTCGTGATGAGCTGCACATTT 400
380 eralagiasp1leuAlaVal1leuengluGluasp1ytrp 396
|||||
401 CCGGCGGAGCATCTCGCGGTCTCTGNAAGGAGGATGGCGGTGG 450
397 ThrVal1uarGasnGlyGlnArgGlyPheValProGlySerTyrrLeuG1 413
|||||
451 ACTGTGAGCGGAGCAACGACGTGCTTGTCCCTTGCTGCTACTTGA 500
413 uLysLeu 415
|||||
501 GAAGCTC 507
seq_name: gb_est45:AM659798

seq_documentation_block: 483 bp mRNA EST 05-APR-2000
LOCUS AM659798
DEFINITION 97502 MARC BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AM659798
VERSION AM659798.1 GI:7425625
KEYWORDS EST
SOURCE Bos taurus.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bovinae; Bos.
1 (bases 1 to 483)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Kohrer,G.A. and
Keeling,W.
Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
Unpublished (2000)
On Jan 6, 2000 this sequence version replaced gi:6675666.
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross-match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAAGCATGACCAT
BACKWARD: GTTTCACATCAGACG
Plate: 71 row: 1 column: 11
Seq primer: ATTAGCTGACATATAG.
Location/Qualifiers
1..483
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCWV SPORT6; Site 1: XbaI; Site 2: XhoI;
library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."

JOURNAL COMMENT
FEATURES
source
BASE COUNT 127 a 126 c 163 g 67 t
ORIGIN
alignment_scores:
Quality: 764.00 Length: 161
Ratio: 4.775 Gaps: 0
Percent Similarity: 99.379 Percent Identity: 90.062
alignment_block:
US-09-068-377-1 x AM659798 ..
Align seg 1/1 to: AM659798 from: 1 to: 483

88 GlySerAlaHis1IleGlnLeuAlaLeuAlaLeuArgGlnLeuArgse 104
|||||
1 GGCACCTCCACATCCAGCTGGCCCTGGCCCTGGCGAGAGCTGGCGAG 50
104 rleuGluGluPheArgGluArgGlnLysGlnArgLysTyrrGlnA 121
|||||
51 CCGGAGGAGTTCCTGAGAGGAGAGAGAGAGAGAGAGAGAGTACGAG 100
121 laIleMetAspArgValGlnLysSerLysLeuSerLeuTyrrLysTr 137
|||||
101 CTGTGATGAGCCGTGTCAGAAAGAGACACTGTGCTCCACAGAGAGCC 150
138 MetGlnSerLysAlaTyrrAspGlnLysCysArgAspAlaAspAla 154
|||||
151 ATGAGATCCAGAGACGATGAGCAGAGAGTCCGGGATGCGGATACGC 200
154 aGluGlnAlaPheGluArgValSerAlaAsnGlyHisGlnLysGlnValG 171
|||||
201 TGAACAGGCGCTTTGAGCGCATAGCACCAACCGTCACAGAGACAGTGG 250
171 LuLysSerGlnAsnLysAlaLysGlnCysLysGlnSerLathrGlnAla 187
|||||
251 AGAAGAGCCAGAACAAAGCCAGACAGTGCAGAGACTCAGCATGAGGCA 300
188 GluArgValTyrrArgGlnAsnIleGlnGlnLeuGluArgAlaArgThrG1 204
|||||
301 GAGCAGACATACAGCAGACAGATGAGCAGCTGGAGAGGCTCGGAGTGA 350
204 uTrrGluGlnGluHisArgThrThrcysGluAlaPheGlnLeuGlnLup 221
|||||
351 GCGGAGCAGGAGCAGACCGCGCACCTGCGCAGCGCTTCCAGTTGCAAGACT 400
221 hAspArgLeuThrIleLeuArgAsnAlaLeuTrrValHisCysAsnG1 237
|||||
401 TTGACCGGCTGACCATCTCCGCCAACTCCCTGTGGGTGACTGCAACGAG 450
238 LeuSerMetGlnCysValLysAspGluLeu 248
|||||
451 CTCCTCATGCACTGCGTCAAGGACGATGAGCTC 483

seq_name: gb_est27:AI952795
seq_documentation_block: 641 bp mRNA EST 08-MAR-2000
LOCUS AI952795
DEFINITION w501f09.x1 NCI CGAP ut1 Homo sapiens cDNA clone IMAGE:2491145 3',
similar to TR:043586 O43586 CD2 BINDING PROTEIN 1 LONG FORM.
;contains Alu repetitive element; mRNA sequence.
ACCESSION AI952795
VERSION AI952795.1 GI:5745105
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 641)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On May 18, 1998 this sequence version replaced gi:3138768.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNU at:
www-bio.llnl.gov/bdrrp/image/image.html
Insert Length: 1978 Std Error: 0.00
Seq primer: -40UP from Gibco

High quality sequence stop: 420.

Location/Qualifiers

source

1. 641

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2491145"

/issue_type="well-differentiated endometrial

adenocarcinoma, 7 pooled tumors"

/lab_host="DH10B"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1; Salt: 1538-014"

Site: 2: Not: Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.75 kb. Life technologies catalog #:

BASE COUNT 103 a 192 c 201 g 143 t 2 others

ORIGIN

alignment_scores:

Quality: 637.50 Length: 160

Ratio: 4.586 Gaps: 1

Percent Similarity: 86.875 Percent Identity: 78.125

alignment_block:

US-09-068-377-1 x AI952795/rev ..

Align seg 1/1 to reverse of: AI952795 from: 1 to: 641

```

257 gluglucysaspvalglugllyaspilleansglypheiieginserysse 273
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
641 GAAGGCTGCACATAGACGACATTCGATCCCGAGCCAGCCAGAG 592
273 rthrgllyarggluprproalaprvalprotyglasntrytyraspa 290
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
591 CACGGGACAGAGCCCTCCGCTCCGCTCCCTACAGAACTATACATC 542
290 rgluvalthrproleuileglyserproserileglinprosercysgly 306
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 GGGAGGTACACCCCGCTACACGACGCCCTGCATACAGCCCTCTGGGC 492
307 valleyllysrpheserglyleuileuileglyserprolystrhrthr 323
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
491 ATGATTAAGAGGTCTCTGACTGCTGCACGGAAGTCCCAAGACCACTTC 442
323 o...SerAlaProAlaSerThrGluThrLeuThrProThrProAla 339
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
441 GTTGCGACGCTTCTGCTGCTGCACAGAGACCTGCACCCCGAGC 392
339 rgaangluvaltyrvalaserilegluvalalinalatrnghnglyasn 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
391 GGAATGAGGGGTGCTACACAGCCATGCAGGAGATACAGGGAAC 342
356 leuasnserialaglnaspytyrargalaleuutyraspytyrthralagl 372
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
341 CCGGCTCTACACGCCGAGGAGTACCGGGCTCTACGATATACAGGCA 292
372 nasenseraspgluleuaspilleansglyaspilleanleuvallel 389
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 GAACCCAGATGAGCTGACCTGCTCCGGGAGACATCTGGAGGTATCC 242
389 euugllygluaspglytyrprthrvalgluargasnlyglnatrgly 405
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 TGGAAAGGGAGGATGGCTGTGACCTGTGAGAGAGGAACGGCGCTGGC 192
406 phevalproglysertryleuclulysleu 415
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
191 TTCGTCCTCCGTGCTCTACCTCGAGAGCTT 162

```

seq_name: gb_est19:AI324866

seq_documentation_block:

LOCUS AI324866 542 bp mRNA EST 23-DEC-1998
DEFINITION m181d02.x1 Soares mouse p3NMF19.5 Mus musculus cDNA clone

IMAGE:472995 3', similar to TR:P97814 P97814 PEST PHOSPHATASE

INTERACTING PROTEIN.; mRNA sequence.

AI324866 GI:4059295

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

The WashU-HIMI Mouse EST Project
Unpublished (1996)
On Jun 15, 1998 this sequence version replaced gi:3227103.
Contact: Marra M/Mouse EST Project
WashU-HIMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:283739
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
Possible reversed clone; similarity on wrong strand
High quality sequence stop: 473.

FEATURES

source

1. 542

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:472995"

/clone_lib="Soares mouse p3NMF19.5"

/dev_stage="19.5 dpc total fetus"

/lab_host="DH10B (ampicillin resistant)"

/note="Vector: pT733 (Pharmacia) with a modified

polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was primed with a Not I - Oligo(dt) primer [5',

TGTTACCATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT733 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by Bento

Soares and M. Fatima Bonaldo. RNA was kindly provided by

Dr. Minoru Ko (Wayne State University)."

BASE COUNT 112 a 152 c 155 g 123 t

ORIGIN

alignment_scores:

Quality: 621.00 Length: 120

Ratio: 5.175 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.167

alignment_block:

US-09-068-377-1 x AI324866/rev ..

Align seg 1/1 to reverse of: AI324866 from: 1 to: 542

```

296 lieglyserproserileglinprosercysglyvalielysrphase 312
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
542 ATTGGCAGCCCTAGGCTCAGCCCTCTGTGTGTGATGAAGAGGTCTTC 493
312 rglyleuileuhtisglyserprolystrthrproseralaproalaaas 329
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
492 TGGGCTGTACTATGAGATCCCAAGACCAACCTTCTCTCTCTGTCTCT 443

```

```

329  erthrGluThrLeuThrProThrProGluArgAsnGluLeuValTyrAla 345
|||||
442  CCACAGAGACTCTGACTCCACCCGAGCGAATGAGTTGCTACGCA 393
|||||
346  SerIleuValGlnAlaThrGlnGlyAsnLeuAsnSerAlaGlnAs 362
|||||
392  TCCATTCACATGCGAGCGAGCCAGGAAACCTTACATCATGACCCAGCA 343
|||||
362  pTYrArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspI 379
|||||
342  CTACCGGGGACTCTACGACTACACCGACAGAAATTCATGATGAGTGA 293
|||||
379  IeserAlaGlyAspIleLeuAlaValIleLeuGlnGlyGluAspGlyTyr 395
|||||
292  TTTCCGGGGAGACATCTCGCGCTCATCTCGAAGGGAGAGATGGCTGG 243
|||||
396  TrrThrValAluArgAsnGlyGlnArgGlyPheValProGlySeryTyr 412
|||||
242  TGGACTGTGGAGCGGAGGACGACACGCTGTGTGCTGCTGGCTACTT 193
|||||
412  uGluTysLeu 415
|||||
192  GGAGAGAGCTC 183

```

seq_name: gb_est43:AM540915

seq_documentation_block:

LOCUS AM540915 547 bp mRNA EST 06-MAR-2000
 DEFINITION C0138F11-3 Mouse E7.5 Extraembryonic Portion cDNA Library Mus
 musculus cDNA clone C0138F11 3, mRNA sequence.

ACCESSION AM540915
 VERSION AM540915.1 GI:7183332

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 547)
 Karzul.G.D., Wang.X., Piao.Y., Dudekula.D., Lim.M.K., Tanaka.T.,
 Sano.Y., Pantano.S., Grahovac.M.J. and Ko.M.S.H.

TITLE Systematic Analyses of NIA Mouse E7.5 Extraembryonic Portion cDNA
 Library

JOURNAL Unpublished (2000)
 COMMENT On Jan 6, 2000 this sequence version replaced gi:6677269.

CONTACT: George J. Karzul
 Laboratory of Genetics

National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@jgsum.gtc.nia.nih.gov

Plate: C0138 row: F column: 11
 Seq primer: -21M13 Forward

High quality sequence stop: 547
 POLY-A=yes.

FEATURES Location/Qualifiers

1..547

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="C0138F11"

/clone_lib="Mouse E7.5 Extraembryonic Portion cDNA
 Library"

/sex="unknown"

/dev_stage="7.5dpc Embryo"

/lab_host="DH10B"

/note="Vector: pSPORT1 (Gibco/BRL Life Technology);
 Site-1: SalI; Site-2: NotI; Total RNAs were extracted from
 5 EPC. The double-stranded cDNA was synthesized by
 Gibco's kit with an Oligo(dt) primer [NotI primer-adaptor
 from GibcoBRL]

[5'-TGACTAGTTCTAGATCGGAGCGGCGGCTTTTCTTTTCTTTT-3']
 from 0.8ug of mRNA. The double-stranded cDNAs were
 treated with T4 DNA polymerase and purified by

ethanol-precipitation. The cDNAs were ligated to

BASE COUNT 114 a 156 c 158 g 119 t

alignment_scores:

Quality: 614.00 Length: 129
 Ratio: 4.797 Gaps: 0
 Percent Similarity: 99.225 Percent Identity: 97.674

alignment_block:

US-09-068-377-1 x AM540915/rev ..

Align seg 1/1 to reverse of: AM540915 from: 1 to: 547

```

287 TYRTYRASPARGLUVALThrProLeuIleGlySerProSerIleGlnP 303
|||||
546 TACTTGTGACAGGAGATG.ACCCACTGATTGGACAGCCCTAGGCTCCAGCC 498
|||||
303 oSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProL 320
|||||
497 CTCCTGTGGTGTATTAAGAGAGTCTCTGGGCTGCTACATGAGATGCCA 448
|||||
320 ySThThrProSerAlaProAlaIaSerThGluThrLeuThrProThr 336
|||||
447 AGACCACACCTTCTGCTCTGCTCCACAGAGACTGTGACTCCACCC 398
|||||
337 ProGluArgAsnGluLeuValTyrAlaSerIleGluValAlaThrGcl 353
|||||
397 CCTGAGCGGATGAGTTGGT.TACGCATTCATCGAGTCAAGCGACGCCA 349
|||||
353 nGlyAsnLeuAsnSerSerAlaGlnAspTyrArgAlaLeuTyrAspTyrT 370
|||||
348 GGGAACCTTAACTCATGACGCCAGGACTACCGGGGACTCTACGACTACA 299
|||||
370 hrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspIleLeuAla 386
|||||
298 CCGCACAGAAATTCATGATGAGCTGACATTTCCGCGGAGACATTCCTGGCG 249
|||||
387 ValIleLeuGluGlyGluAspGlyTrrPThrValGluArgAsnGlyG 403
|||||
248 GTGATCTCTGGAAAGGGAGAGATGGCTGGTGGACTTTGGACCGGAGACGACA 199
|||||
403 nArgGlyPheValProGlySeryTyrLeuGluTysLeu 415
|||||
198 ACGTGGCTTTGTCCTGGTCTACTTGGAGAGAGCTC 162

```

seq_name: gb_est42:AM407004

seq_documentation_block:

LOCUS AM407004 396 bp mRNA EST 16-FEB-2000
 DEFINITION UI-HE-BL0-af-d-08-0-UI.r1 NIH-MGC_37 Homo sapiens cDNA clone
 IMAGE:3061310 5', mRNA sequence.

ACCESSION AM407004
 VERSION AM407004.1 GI:6926061

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 396)
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT On May 20, 1999 this sequence version replaced gi:4880298.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov

Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILM at: www.bio.lnl.gov/bbrp/image/image.html
Seq primer: M3 Forward.

FEATURES

source
1..396
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3061310"
/clone_lib="NIH-MGC_37"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/lab_host="DH10B (LTI)"
/note="Vector: pT73-Pac; Site_1: NotI; Site_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (1.5-2.5kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT
95 a 130 c 111 g 60 t

alignment_scores:
Quality: 531.50 Length: 131
Ratio: 4.504 Gaps: 1
Percent Similarity: 90.076 Percent Identity: 78.626

alignment_block:

US-09-068-377-1 x AM407004 ..

Align seg 1/1 to: AM407004 from: 1 to: 396

```

246 ASpgluuEuTYrgluGluValArgLeuThrleuGluGlyCysaspValG1 262
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
      4 GAGGAGCTCTACGAGAGAGTCCGCTGAGAGGCTGACACATAGA 53
262 uGlyAspIleasnGlyPheIleGlnSerLysSerThrGlyArgGluProp 279
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
      54 CGCGACATCGACAGTTCATCCAGGCCAAGACAGCAGGCGACAGGCC 103
279 roAlaProValProTYrGlnAsnTYrTrpAspArgGluValAlaThrProLeu 295
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
      104 CCGCTCGGCTGCTTACCAAACTATTACATGCGAGGAGTCCACCCGCTG 153
296 IlleGlySerProSerIleGlnProSerCysGlyValIleLysArgPhe 312
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
      154 ACCAGCAGCCCTGGCATACAGCCGCTCTGTGGCATATAAGAGTTCTC 203
312 rGlyLeuLeuHisGlySerProLysThrThrPro...SerAlaProAla 328
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
      204 TGGACTGCTGACAGAGTCCCAAGACACTTCGTTGCGAGCTTCGCTG 253
328 laSerThrGluThrLeuThrProThrProGluArgAsnGluLeuValTYr 344
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
      254 CGTCCACAGAGACCTGACCCACCCCGAGCGGAAATGAGGCTGTCTAC 303
345 AlaSerIleGluValGlnAlaThrGlnGlyAsnLeuAsnSerSerAlaG1 361
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
      304 ACAGCCATCGCAGTGCAGAGATACAGGAACCCGCGCTCACACGCCA 353
361 nAspTYrArgAlaLeuTYrAspTYrThrAlaGlnAsnSerasp 375

```

seq_name: gb_est45:AM659454

354 GGAGTACCGCGGCTTACGATTATACAGCGAGAACCCAGAT 396

seq_documentation_block:

LOCUS AM659454 366 bp mRNA EST 05-APR-2000
DEFINITION 96640 MARC 1BOV Bos taurus CDNA 5', mRNA sequence.
ACCESSION AM659454
VERSION AM659454.1 GI:7425281
KEYWORDS EST.
SOURCE Bos taurus.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovine; Bos.
1 (bases 1 to 366)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
Keeler,J.W.
Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
Unpublished (2000)
On Jan 6, 2000 this sequence version replaced gi:6675298.
Contact: Smith TP.
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross-match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCAGAC
Plate: 74 row: A column: 5
Seq primer: ATTAGTGACATATAG.

FEATURES

source
1..366
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."

alignment_scores:
Quality: 516.00 Length: 105
Ratio: 4.962 Gaps: 0
Percent Similarity: 99.048 Percent Identity: 88.571

alignment_block:

US-09-068-377-1 x AM659454 ..

Align seg 1/1 to: AM659454 from: 1 to: 366

```

188 GluArgValTYrArgGlnAsnIleGluGlnLeuGluArgAlaArgThrG1 204
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
      48 GAGCGACATACAGAGACATTTGACAGCTGAGAGGGGTGCGAGTGA 97
204 uTrrGluGlnGluHisArgThrThrCysGluAlaPheGlnLeuGlnGluP 221
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
      98 GTGGGACAGAGAGACCCGCCACCTCGAGGCTTCAGTTGACAGAGT 147
221 heAspArgLeuThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGln 237
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
      148 TTGACCGCTGACCATCTCCGCAACTCCTGTGGGTGACATGCAACAG 197

```

238 LeuSerMetGlnCysValIysAspAspGluLeuTyrGluGluValArgLe 254
 |||||||
 198 CTCCTCATGAGAGTGCCTCAAGAGCATGAGCTCTACGAAAGAGTGGGGCT 247
 |||||||
 254 uThrLeuGluGlyCysAspValGluGlyAspIleAsnGlyPheIleGlns 271
 :|||||
 248 GAGCTGGAGGCTGAGCGGTGGAGCGGACATTGACGGCTTCATCCAGG 297
 |||||||
 271 eTyrSerThrGlyArgGluProProAlaProValProTyrGlnAsnTyr 287
 :|||||
 298 CCAAGACACAGCGGACACTGAGGCCCGACGCCAGTCCCTTACCAGAACTAC 347
 |||||||
 288 TyrAspArgGluVal 292
 |||||||
 348 TATGATCGGAGGTG 362

seq_name: gb_est13:AA893267

seq_documentation_block: 472 bp mRNA EST 16-JUN-1998
 LOCUS AA893267
 DEFINITION EST197070 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
 R1BE29 3' end, mRNA sequence.
 ACCESSION AA893267
 VERSION AA893267.1 GI:3020146
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 472)
 Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
 Kerlavage,A.R. and Adams,M.D.
 Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
 Gene Index
 Unpublished (1998)
 On Jan 5, 1998 this sequence version replaced gi:2581190.
 COMMENT
 Contact: Lee, NH
 ATCC
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@ligr.org
 Seq primer: M13-21.

FEATURES Location/Qualifiers

1..472
 /organism="Rattus sp."
 /db_xref="ATCC (inhost):2014762"
 /db_xref="taxon:10118"
 /clone="R1BE29"
 /clone_lib="Normalized rat kidney, Bento Soares"
 /note="Organ: Kidney; Vector: pT7T3Pac; Site_1: EcoRI;
 Site_2: NotI"
 BASE COUNT 99 a 133 c 136 g 104 t
 ORIGIN

alignment_scores:
 Quality: 501.50 Length: 105
 Ratio: 4.869 Gaps: 1
 Percent Similarity: 98.095 Percent Identity: 94.286

alignment_block:

US-09-068-377-1 x AA893267/rev ..

Align seg 1/1 to reverse of: AA893267 from: 1 to: 472

312 SerGlyLeuLeuHisGlySerProLysThrTherProSer...AlaProAl 327
 |||||||
 470 TCTGGGCTGCTACATGGAAGTCCCAAGACCAACACCTTCTCAAGCTCTGC 421

327 aAlaSerThrGlnThrLeuThrProThrProGluArgAsnGluLeuValT 344
 |||||||
 420 CGCTTCACAGAGACTGTGGCTCCACCTCTGAGCGAAGAAATGTGCT 371
 |||||||
 344 yTAlaSerIleGluValGlnAlaThrGlnGlyAsnLeuAsnSerAla 360
 :|||||
 370 AGCATCCATTGAAAGTGCAGGACCAACCCAGCGGACCTTAACACAGACC 321
 |||||||
 361 GlnAspTyrArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGlu 377
 :|||||
 320 CAGGACTACCGGGCCCTCTATGACTACACGCGACGAATTCGATGAGCT 271
 |||||||
 377 uAspIleSerAlaGlyAspIleLeuAlaValIleLeuGluGlyGluAspG 394
 |||||||
 270 GGACATTCTTCTGGCGGAGACATCTCGCAGTCAATCTGGAAGGAGGATG 221
 |||||||
 394 lYTrpTrpThrValGluArgAsnGlyGlnArgGlyPheValProGlySer 410
 |||||||
 220 GCTGGTGACTGCTGGAGCAACGACCAACCTGTGCTGCTGCTGCTGCTG 171
 |||||||

seq_name: gb_est14:AA980931

seq_documentation_block: 361 bp mRNA EST 27-MAY-1998
 LOCUS AA980931
 DEFINITION ua47e05.t1 Soares_mammary-gland_NbMWG Mus musculus cDNA clone
 IMAGE:1349888 5' similar to tr:p97814 p97814 PEST PHOSPHATASE
 INTERACTING PROTEIN. ;, mRNA sequence.
 ACCESSION AA980931
 VERSION AA980931.1 GI:3159467
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 361)
 Geisel,S., Kucaba,T., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 On Apr 7, 1998 this sequence version replaced gi:3036147.
 COMMENT
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LINT ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:698680
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 276.

FEATURES source

1..361
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:1349888"
 /clone_lib="Soares_mammary-gland_NbMWG"
 /sex="male"
 /tissue_type="mammary gland"
 /dev stage="4 weeks"
 /lab_host="DH10B"
 /note="Organ: mammary gland; Vector: pT7T3b-Pac
 (Pharmacia) with a modified polylinker; Site_1: Not I;
 Site_2: Eco RI; 1st strand cDNA was primed with a Not I -

oligo(dT) primer [5'
TGTTCACCATCTGAAGTGGAGCGCCGCCGAAGTGTGTGTGTGTGTGTGTGTGT
T 3'] double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."

BASE COUNT 87 a 90 c 108 g 76 t
ORIGIN
alignment_scores:
Quality: 488.50 Length: 107
Ratio: 4.985 Gaps: 1
Percent Similarity: 91.589 Percent Identity: 87.850

alignment_block:
US-09-068-377-1 x AA980931 ..
Align seg 1/1 to: AA980931 from: 1 to: 361

237 GlnLeuSerMetGlnCysValIysAspGluLeuTyrGluGluValAr 253
|||||
1 CAGCTCTCCATGCATGTGTCAAGATGATGACCTTATGAGAGAGTCGC 50
253 GlnThrLeuGluGlyCysAspValIglIysAspIleasnGlyPheIleG 270
|||||
51 GCTGACCTTGAGGGCTGTGATGTGAAGGTGACATCATGGCTTCATCC 100
270 InSerIleSerThrGlyArgGluProProAlaProValProTyrGlnAsn 286
|||||
101 AGTCGAAAGACCTGGCAGAGAGCCGCCAGCTCCGCTTATCAGAAC 150
287 TyrTyrAspArgGluValThrProLeuIleGlySerProSerIleGlnPr 303
|||||
151 TACTATGACAGGAGAGTGCACCCACTGATGTGCAGCCCTACATCCAGCC 200
303 oSerCysGlyValIleIysArgPheSerGlyLeuLeuHisGlySerProL 320
|||||
201 CTCTGGCGGTGTGATAAAGAGGTCTCTGGGCTCTCATGGAAGTCCCA 250
320 ysrThrProSerAlaProAlaAla.....Ser 329
|||||
251 AGACACACACCTTCCTGCTCTCTCTGTAATAATAGGGGAGAGGCTTAATGG 300
330 ThrGluThrLeuThrProThr 336
|||||
301 ACAGGGAGACATGTCTCCACCC 321
seq_name: gb_est45:AW657869

seq_documentation_block:
LOCUS AW657869 572 bp mRNA EST 05-APR-2000
DEFINITION 93496 MARC 1BOV Bos taurus CDNA 5', mRNA sequence.
ACCESSION AW657869
VERSION AW657869.1 GI:7423695
KEYWORDS EST
SOURCE Bos taurus.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 572)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
Keefe,J.W.
Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.960904.e. Vector identified by cross_match with the -mismatch 18
and -mismatch 12 options.
PCR primers
FORWARD: AGGAACACCTATGACCAT
BACKWARD: GTTTCACGATCAGCAGC
Plate: 79 row: C column: 7
Seq primer: ATTAGTGCATCTATAG.
Location/Qualifiers
1.572
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."

FEATURES
source
location/Qualifiers
1.572

BASE COUNT 139 a 142 c 186 g 105 t
ORIGIN

alignment_scores:
Quality: 456.00 Length: 102
Ratio: 4.515 Gaps: 0
Percent Similarity: 99.020 Percent Identity: 88.235

alignment_block:
US-09-068-377-1 x AW657869 ..

Align seg 1/1 to: AW657869 from: 1 to: 572
71 AsnSerLeuArgThrSerPheAspSerLeuIleGlnGlnThrGluAsnVa 87
:::|||||
267 AGCTCCCTGAGCGCCCTCTTGACTCTGACAGCAGCAAGATGAGATGT 316
87 IglSerAlaHisIleGlnLeuAlaLeuAlaLeuArgGluGluLeuArgS 104
|||||
317 GGGAGCTCCACATCCAGCTGGCCCTGGCCCTGGCCGAGAGAGCTCGCA 366
104 erLeuGluGluPheArgGluArgGlnIleGluGlnIleArgIleGlyIle 120
|||||
367 GCTGAGAGAGTTCCTGAGAGGACAGAGAGAGAGAGAGAGAGAGAGAG 416
121 AlaIleMetAspArgValGlnIleSerIleLeuSerLeuTyrIleIle 137
|||||
417 GCTGTATGAGACCGTGTCCAGAGAGCAAGCTGTCTCCACAGAGAGGC 466
137 rMetGluSerIleValAlaTyrAspGlnIleIleIleIleIleIleIleIle 154
:|||||
467 CATGAGACTCCAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 516
154 IacGluAlaIlePheGluArgValIleSerAlaAsnGlyHisGlnIleGlnVal 170
|||||
517 CTGAACAGAGCTTTGAGCGCATTTAGACACCAAGCTCAACAGAGAGAGTG 566
171 GluLys 172
|||||
567 GAGAAAG 572

seq_name: gb_est45:AW654760
seq_documentation_block:
LOCUS AW654760 543 bp mRNA EST 05-APR-2000
DEFINITION 104853 MARC 1BOV Bos taurus CDNA 5', mRNA sequence.
ACCESSION AW654760
VERSION AW654760.1 GI:7420586
KEYWORDS EST
SOURCE Bos taurus.
ORGANISM Bos taurus

REFERENCE
AUTHORS

1 (bases 1 to 543)
Smith, T.P.L., Csaacs, E., Stone, R.T., Heaton, M.P., Grosse, W.M.,
Bovidae, Bovinae: Bos.
Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Eumariota: Euteria: Cetartiodactyla: Ruminantia: Pecora: Bovidae:
Bovidae, Bovinae: Bos.

TITLE	Design and use of four pooled tissue normalized cDNA libraries for
JOURNAL	EST discovery in cattle
COMMENT	Unpublished (2000)
	On Oct 8, 1998 this sequence version replaced g1:3729921.

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -mismatches
and -minmatch 12 options.

PCR primers
FORWARD: AGGAACAGCTATGACCACAT
BACKWARD: GTTTCACAGTCACGACG
Plate: 100 row: D column: 4
Seq primer: ATTAGCTGACACTATAC.

386 AlValIIleengluGIyIuIuspsGIytrpItrpItrpIValGIuaGaaNGI 402
336 GAGGTATCTCTGGAAGGAGGAGGACGGCTGTGTGACAGGTACAAACGAAACGG 385
402 yGIuaNglyIphValaProGlySerTyrIleuGIuIusIeu 415
386 GCAGCGGTGGCTTGTGCTCTCTCTACTTGGAGAAAGCTT 425

seq_name: gb_est45:AW654759

```
seq_documentation_block:
```

TITLE	Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle
JOURNAL	Unpublished (2000)
COMMENT	On May 18, 1998 this sequence version replaced g1:3137243.

Email: smthe@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -mismatch 12 options.
PCR primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGCTACGACG
Plate: 100 row: D column: 2
Seq primer: ATTAGGTGACACTATAG.

```
Align seg 1/1   to: AW654759    from: 1 to: 553  
287 YYYTYRASPATAGGLVAlThrProLeuIleGlySerProSerIleGlnPr 303  
||||| :|||:  
66 TACTTCGCCCGCCGTCACTCACACCTCCTTCTGCCTCCCC..... 104A  
303 oserysglyValIIeUysargPheSerGlyLeuLeuHisIcylserProL 322E  
|||||
```

```

105 .....AGCTTTCTGGGCTGCTGCATGAGATGCCA 135
320 ysthrProSer...AlaProAlaAlaSerThrGluThrLeuThrPro 335
136 AGACACACAGTGTGACAGCTTCTGCACCTCCACAGAGACCTGACTCT 165
336 ThrProGluArgAsnGluLeuValTyrAlaSerIleGluValAlaGln 352
186 ACCCTTACCGAGATGAGGTGTCTACGCTGCATGCGCTGCAGGAGGC 235
352 rGlnGlyAsnLeuAsnSerSerAlaGlnAspTyrArgAlaLeuTyrAsp 369
236 GCCAGGGCCCCCACCACATGCCAGGCCAGCACTATAGCCACTATGACT 285
369 ythrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspIleLeu 385
286 AACACACCCCAAGAACTGTGACAGCTGACATCTCTCGGGGACATCTG 335
386 AlaValIleLeuGluGlyGluAspGlyTyrTrpThrValGluArgAsn 402
336 GAGGTATCTGTGAAAGGAGGAGGAGCTGTGTGACGCTACAAAGGAA 385
402 yGlnArgGlyPheValProGlySerTyrLeuGluIleu 415
386 GCAGCTGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 425
seq_name: gb_est13:AA900024

seq_documentation_block:
LOCUS AA900024 420 bp mRNA EST 04-JUL-1999
DEFINITION UI-R-E0-dh-01-0-01.s1 UI-R-E0 Rattus norvegicus cDNA clone
UI-R-E0-dh-01-0-01.3, similar to g11857711|gb|U87814|MM087814
Mus musculus PEST phosphatase interacting protein mRNA, complete
cds, mRNA sequence.
ACCESSION AA900024
VERSION AA900024.1 GI:4232513
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 420)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Apr 7, 1998 this sequence version replaced gi:3035378.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@iuii.wiwi.iowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult 8-day-embryo library. cDNA library Preparation: M. Fatima
Bonaldo, Ph.D. Clone distribution: clones will be available through
Research Genetics This clone is also available through the
I.M.A.G.E. Consortium at LNL (info@image.llnl.gov). IMAGE
ID-1770653
Seq primer: M13 Forward
POLYA-No.
FEATURES
Source
Location/Qualifiers
1..420
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="UI-R-E0-dh-01-0-01"
/dev_stage="embryonic"

```

```

/ab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site-1: NotI; Site-2: EcoRI. This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture."
BASE COUNT 84 a 120 c 107 g 106 t 3 others
ORIGIN

alignment_scores:
Quality: 400.00 Length: 81
Ratio: 5.000 Gaps: 0
Percent Similarity: 98.765 Percent Identity: 95.062

alignment_block:
US-09-068-377-1 x AA900024/rev ..

Align seg 1/1 to reverse of: AA900024 from: 1 to: 420
335 ProThrProGluArgAsnGluLeuValTyrAlaSerIleGluValAlaGln 351
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
415 CCCACTWTGAGCAGAAAGAAATGTTATGCAATCCTTAAGTGCAGGC 366
351 aThrGlnGlyAsnLeuAsnSerSerAlaGlnAspTyrArgAlaLeuTyr 368
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
365 AACACAGGGCAACCTTAACACAGAGCCAGGACTACCGGGCCTGATG 316
368 sPyrThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspIle 384
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315 ACACACAGGCAAGAAATTCATGATGAGTGCACATTCTGGGAGACATC 266
385 LeuAlaValIleLeuGluGlyGluAspGlyTyrTrpThrValGluArgAs 401
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
265 CTGCACTCATCTCTGAGAGGAGATGCTGTGAGACGTGAGCAAGAA 216
401 nGlnArgGlyPheValProGlySerTyrLeuGluIleu 415
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
215 CGGACAACGCGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 173
seq_name: gb_est54:w85898

seq_documentation_block:
LOCUS w85898 429 bp mRNA EST 02-FEB-1997
DEFINITION zh56907.s1 Soares fetal liver_spleen_INFIS.S1 Homo sapiens cDNA
clone IMAGE:416124.3, similar to SW:MTSC_ACACA P10569 MYOSIN IC
HEAVY CHAIN. [1];, mRNA sequence.
ACCESSION w85898
VERSION w85898.1 GI:1398327
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 429)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapelli, B.,
Chissoe, S., Dietrich, N., Dubugue, T., Favello, A., Gish, M.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Trevasakis, E., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
JOURNAL genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT On Nov 6, 1997 this sequence version replaced gi:933404.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

```

Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through INL; contact the
IMAGE Consortium (imageimage.lnl.gov) for further information.
Insert Length: 448 Std Error: 0.00
Seq primer: mob.RGA+RT
High quality sequence stop: 375.

FEATURES

source

```

1. 429
/organism="Homo sapiens"
/db_xref="GB:1324594"
/db_xref="taxon:9606"
/clone="IMAGE:416124"
/clone_lib="Soares_fetal_liver_spleen_INFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3 (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5',
AAGTGAGAAATTAATTAAGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Patricia Bonaldo."
BASE COUNT      67 a   132 c   136 g   91 t      3 others
ORIGIN

```

alignment_scores:
Quality: 393.00 Length: 91
Ratio: 4.852 Gaps: 0
Percent Similarity: 89.011 Percent Identity: 84.615

alignment_block:
US-09-068-377-1 x W85898/rev ..

Align seg 1/1 to reverse of: W85898 from: 1 to: 429

```

325 AlapProAlaAaserThrGluThrLeuThrProThrProGluArgAsnG1 341
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
405 GCTTCTGCTGCTGCACAGACCCCTGACCCCGGCGGGAATGA 356
341 uLeuValTyrAlaSerTleGluValGlnAlaThrGlnGlyAsnLeuAsnS 358
| ||||| : : : : : ||||| ||||| ||||| |||||
355 GGCTGTCTACACAGCCATCGCAGTGCAGAGATACAGGAAACCCGGGCTT 306
358 eTserAlaGlnAspTyrArgAlaLeuTyrAspTyrThrAlaGlnAsnSer 374
|| ||||| : : : : : ||||| ||||| ||||| |||||
305 CACGAGCCCGAGATACCGGGCGCTACGATTTACAGCGCAGAACCCA 256
375 AspGluLeuAspTleSerAlaGlyAspTleLeuAlaValIleLeuGluG1 391
||||| : : : : : ||||| ||||| ||||| |||||
255 GATGAGCTGAGCTGTCCGGGAGACATCTGAGGTGATCTGGAAGG 206
391 yGluAspGlyTyrThrPheValGluArgAsnGlyGlnArgGlyPheValP 408
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
205 GGAAGATGGCTGGTGAGCTGTGAGAGAAAGGACGCGCTTCGCTCC 156
408 roGlySerTyrLeuGluIuysLeu 415
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
155 CTGCTCTCTACTGAGAGACTT 133

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2000, 10:37:38 ; Search time 48.19 Seconds

(without alignments)
597.088 Million cell updates/sec

Title: US-09-068-377-1

Perfect score: 2152
Sequence: 1 MAAQLQFDFARWCGRDFTAH.....WTVERNGRGFVPSYLEKL 415

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 segs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SP:archaea:*
2: SP:bacteria:*
3: SP:fungi:*
4: SP:human:*
5: SP:invertebrate:*
6: SP:mammal:*
7: SP:mhc:*
8: SP:organelle:*
9: SP:phage:*
10: SP:plant:*
11: SP:rodent:*
12: SP:virus:*
13: SP:vertebrate:*
14: SP:unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2152	100.0	415	11	P97814
2	1913.5	88.9	416	4	O43586
3	1877.5	87.2	416	4	O95657
4	1748	81.2	397	4	O43585
5	736.5	34.2	334	11	O92189
6	283.5	13.2	486	11	O9WVE8
7	283	13.2	906	4	O14526
8	266.5	12.4	441	11	O920W5
9	266.5	12.4	448	13	O13154
10	265.5	12.3	441	11	O61644
11	238.5	11.1	336	4	O60861
12	238.5	11.1	412	4	O43144
13	236.5	11.0	421	11	O60780
14	227	10.5	441	11	O55148
15	206	9.6	1146	11	O9WVE1
16	206	9.6	1217	11	O9WVE9
17	202	9.4	1721	4	O95216
18	199.5	9.3	1011	5	O61639
19	199.5	9.3	1094	5	O61618

20	198	9.2	1714	11	O920R4	O920R4 mus musculu
21	194.5	8.9	1270	13	O42287	O42287 xenopus lae
22	191.5	8.9	416	5	O07839	O07839 echinococcu
23	188.5	8.6	1658	11	O920R6	O920R6 mus musculu
24	187.5	8.7	674	4	O60301	O60301 mus musculu
25	187	8.7	1197	11	O920R5	O920R5 mus sapien
26	186.5	8.6	547	11	P97531	P97531 rattus norv
27	185.5	8.7	2427	5	O21408	O21408 caenorhabdi
28	183	8.5	951	5	O16647	O16647 caenorhabdi
29	179.5	8.3	433	11	O62418	O62418 mus musculu
30	178.5	8.3	471	11	O70419	O70419 rattus norv
31	177.5	8.2	426	5	O07840	O07840 echinococcu
32	175.5	8.2	335	4	O9Y4V2	O9Y4V2 homo sapien
33	175.5	8.2	545	4	O15184	O15184 homo sapien
34	171.5	8.0	508	11	O70420	O70420 rattus norv
35	171	7.9	1186	5	O61080	O61080 acanthameb
36	170	7.9	785	5	O9XU57	O9XU57 caenorhabdi
37	169.5	7.9	434	11	O62434	O62434 mus musculu
38	169	7.9	1099	13	O90748	O90748 gallus gall
39	167	7.8	684	11	O62417	O62417 mus musculu
40	167	7.8	714	11	O92028	O92028 mus musculu
41	167	7.8	724	11	O92029	O92029 mus musculu
42	166.5	7.7	1736	4	O75111	O75111 homo sapien
43	166.5	7.7	1857	4	O95153	O95153 homo sapien
44	165	7.7	684	4	O9Y338	O9Y338 homo sapien
45	164.5	7.6	1014	4	O76041	O76041 homo sapien

ALIGNMENTS

RESULT 1
ID P97814 PRELIMINARY; PRT: 415 AA.
AC P97814;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE PEST PHOSPHATASE INTERACTING PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA LASKY L.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U07814; AAB48483.1; -.
DR HSSP; P07751; IRL2.
DR PFAM; PF00611; FCH; 1.
DR PFAM; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
SQ SEQUENCE 415 AA; 47590 MW; C1C5FB4C CRC32;

Query Match 100.0%; Score 2152; DB 11; Length 415;
Best Local Similarity 100.0%; Pred. No. 3.1e-131;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAQLQFDFARWCGRDFTAHGTGEVLLQRLDGRKKCKYVEELLRQRAQAEERYGGEIYVI 60
DB 1 MAAQLQFDFARWCGRDFTAHGTGEVLLQRLDGRKKCKYVEELLRQRAQAEERYGGEIYVI 60
QY 61 ARKAGGCGEENSLKRTSPSLKOOFTENVSANHTOTALAREELRSLEEFREKQRRKYE 120
DB 61 ARKAGGCGEENSLKRTSPSLKOOFTENVSANHTOTALAREELRSLEEFREKQRRKYE 120
QY 121 AIMDRVOKSKLSLYKKTWESKRAYDOKCRDADDAQAFERYSANGHQVKSQKAKQC 180
DB 121 AIMDRVOKSKLSLYKKTWESKRAYDOKCRDADDAQAFERYSANGHQVKSQKAKQC 180
QY 181 KESATEARVYRONIEOLERARTEDQHRRTTCGAFQLOEFDRLITLNALMVHGNQLSM 240
DB 181 KESATEARVYRONIEOLERARTEDQHRRTTCGAFQLOEFDRLITLNALMVHGNQLSM 240

QY 241 QCVADDELYEVRRLTEGCDVEGDIINGFIQSKSTGREPPAPVPONYDYREVTPLIGSPS 300
|||||
Db 241 QCVADDELYEVRRLTEGCDVEGDIINGFIQSKSTGREPPAPVPONYDYREVTPLIGSPS 300
QY 301 IQPSCGVYKRRSGLLHSGPKTTPSPAPASTETLTPPERNELVYASIEVQATOGNLNSA 360
|||||
Db 301 IQPSCGVYKRRSGLLHSGPKTTPSPAPASTETLTPPERNELVYASIEVQATOGNLNSA 360
QY 361 QDYRALDYTAQNSDELISAGDILAVILLEGEDGWMYVERNORGFPVGSYLEKL 415
|||||
Db 361 QDYRALDYTAQNSDELISAGDILAVILLEGEDGWMYVERNORGFPVGSYLEKL 415

RESULT 2

043586 PRELIMINARY; PRT; 416 AA.
AC 043586;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CD2 BINDING PROTEIN 1 LONG FORM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA LI J., NASHIZAMA K., AN W., HUSSEY R.E., LIALIOS F.E.,
RA SUNDER-PLESSMAN R., REINHERZ E.L.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF038603; AADI1959.1; -.
DR HSSP; P07751; IBK2.
DR PFAM; PF00611; FCH; 1.
DR PFAM; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
SQ SEQUENCE 416 AA; 47591 MW; 9EBAB899 CRC32;

Query Match 88.9%; Score 1913.5; DB 4; Length 416;
Best Local Similarity 88.0%; Pred. No. 6.8e-116;
Matches 366; Conservative 20; Mismatches 29; Indels 1; Gaps 1;

QY 1 MMAOLFQDAFWCNDFTAHNGYEVLQRLDGRKMKDVEELLRQRAQAEERYGKELVQI 60
|||
Db 1 MMPQLQKDAFWCNDFTAHNGYEVLQRLDGRKMKDVEELLRQRAQAEERYGKELVQI 60
QY 61 ARKAGGOTENSLRTSPDSLKQOTENGSAAHIQALALREELRSLEPEREKQORRKYE 120
|||||
Db 61 ARKAGGOTENSLRTSPDSLKQOTENGSAAHIQALALREELRSLEPEREKQORRKYE 120
QY 121 AIMDRVOKSKLSLYKKTMESKKAAYDQRCRDADAEOAFERYVANGHOKYKESONKAKOC 180
|||||
Db 121 AIMDRVOKSKLSLYKKTMESKKAAYDQRCRDADAEOAFERYVANGHOKYKESONKAKOC 180
QY 181 KESATEAERYRONIEQLEKARTEMOEHRITCEAFQLOEFDRLTILRNALMWHNCOLSM 240
|||||
Db 181 KESATEAERYRONIEQLEKARTEMOEHRITCEAFQLOEFDRLTILRNALMWHNCOLSM 240
QY 241 QCVADDELYEVRRLTEGCDVEGDIINGFIQSKSTGREPPAPVPONYDYREVTPLIGSPS 300
|||||
Db 241 QCVADDELYEVRRLTEGCDVEGDIINGFIQSKSTGREPPAPVPONYDYREVTPLIGSPS 300
QY 301 IQPSCGVYKRRSGLLHSGPKTTPSPAPASTETLTPPERNELVYASIEVQATOGNLNSA 359
|||||
Db 301 IQPSCGVYKRRSGLLHSGPKTTPSPAPASTETLTPPERNELVYASIEVQATOGNLNSA 359
QY 360 AODYRALDYTAQNSDELISAGDILAVILLEGEDGWMYVERNORGFPVGSYLEKL 415
|||||
Db 360 AODYRALDYTAQNSDELISAGDILAVILLEGEDGWMYVERNORGFPVGSYLEKL 415

RESULT 3
095657

ID 095657 PRELIMINARY; PRT; 416 AA.
AC 095657;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PEST PHOSPHATASE INTERACTING PROTEIN HOMOLOG.
GN H-PIP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA WILSON L.A., FIELDS D., CRUZ L., LASKY L., FRIESEN J.,
RA SIMONOVITCH K.A.;
RT "The human homologue of mouse ppp-PIP interactor protein."
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U94778; AAD00762.1; -.
DR HSSP; P07751; IBK2.
SQ SEQUENCE 416 AA; 47610 MW; F866F424 CRC32;

Query Match 87.2%; Score 1877.5; DB 4; Length 416;
Best Local Similarity 86.1%; Pred. No. 1.4e-113;
Matches 358; Conservative 23; Mismatches 34; Indels 1; Gaps 1;

QY 1 MMAOLFQDAFWCNDFTAHNGYEVLQRLDGRKMKDVEELLRQRAQAEERYGKELVQI 60
|||
Db 1 MMPQLQKDAFWCNDFTAHNGYEVLQRLDGRKMKDVEELLRQRAQAEERYGKELVQI 60
QY 61 ARKAGGOTENSLRTSPDSLKQOTENGSAAHIQALALREELRSLEPEREKQORRKYE 120
|||||
Db 61 ARKAGGOTENSLRTSPDSLKQOTENGSAAHIQALALREELRSLEPEREKQORRKYE 120
QY 121 AIMDRVOKSKLSLYKKTMESKKAAYDQRCRDADAEOAFERYVANGHOKYKESONKAKOC 180
|||||
Db 121 AIMDRVOKSKLSLYKKTMESKKAAYDQRCRDADAEOAFERYVANGHOKYKESONKAKOC 180
QY 181 KESATEAERYRONIEQLEKARTEMOEHRITCEAFQLOEFDRLTILRNALMWHNCOLSM 240
|||||
Db 181 KESATEAERYRONIEQLEKARTEMOEHRITCEAFQLOEFDRLTILRNALMWHNCOLSM 240
QY 241 QCVADDELYEVRRLTEGCDVEGDIINGFIQSKSTGREPPAPVPONYDYREVTPLIGSPS 300
|||||
Db 241 QCVADDELYEVRRLTEGCDVEGDIINGFIQSKSTGREPPAPVPONYDYREVTPLIGSPS 300
QY 301 IQPSCGVYKRRSGLLHSGPKTTPSPAPASTETLTPPERNELVYASIEVQATOGNLNSA 359
|||||
Db 301 IQPSCGVYKRRSGLLHSGPKTTPSPAPASTETLTPPERNELVYASIEVQATOGNLNSA 359
QY 360 AODYRALDYTAQNSDELISAGDILAVILLEGEDGWMYVERNORGFPVGSYLEKL 415
|||||
Db 360 AODYRALDYTAQNSDELISAGDILAVILLEGEDGWMYVERNORGFPVGSYLEKL 415

RESULT 4

043585 PRELIMINARY; PRT; 397 AA.
AC 043585;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CD2 BINDING PROTEIN 1 SHORT FORM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA LI J., NASHIZAMA K., AN W., HUSSEY R.E., LIALIOS F.E.,
RA SUNDER-PLESSMAN R., REINHERZ E.L.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF038602; AADI1958.1; -.
DR HSSP; P07751; IBK2.
DR PFAM; PF00611; FCH; 1.


```

Db 354 STQLOSSNPFDEDDTSSSISEKEDIKANKVSSYEKTQYPTDWSDDSNPFSSDAN 413
QY 356 LNNSAOD-----RATLYDYTAONSDELISAGDILVILEGEGWMTVER--NGO 403
Db 414 GDSNPFDDDTTSGTFVRKALYDYGQSHDELSTFRAGDELTKIEDEDEQGWCKGRLDSDG 473
QY 404 RGFVPGSTLEKL 415
Db 474 VGLYPANYVEAI 485

```

```

RESULT 7
ID 014526 PRELIMINARY; PRT; 906 AA.
AC 014526;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-AUG-1999 (Tremblrel. 11, Last annotation update)
DE KIA0290 (FRAGMENT).
GN KIA0290.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA OHARA O., NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N.,
RA NOMURA N.;
RN Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE OF 662-906 FROM N.A.
RA LAMERDIN J.E., MCCREADY P.M., SKORONSKI E., VISWANATHAN V.,
RA BURKHART-SCHULTZ K., GORDON L., DIAS J., RAMIREZ M., STILWAGEN S.,
RA PHAN H., VELASCO N., DO L., REGALA W., TERRY A., GARNES J.,
RA DANGNAN L., EBLER A., CHRISTENSEN M., GEORGESE A., AVILA J., LIU S.,
RA ATTIX C., ANDRIESE T., FRANKHEIM M., AMICO-KELLER G., COFFIELD J.,
RA DUARTE S., LUCAS S., BRUCE R., THOMAS P., QUAN G., KROMMILLER B.,
RA ARELLANO A., SAUNDERS C., OW D., NOLAN M., TRONG S., KOBAYASHI A.,
RA OLSEN A.S., CARRANO A.V.;
RN "Sequence analysis of an ~700 kb region in 19p13.1 between JAK3 and
RT PDE4C."
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB006628; BAA22959.1; -;
DR EMBL; AC005952; AAC72276.1; -;
FT NON-TER
SQ SEQUENCE 906 AA; 98738 MW; 61E95187 CRC32;

```

```

Query Match 13.2% Score 283; DB 4; Length 906;
Best local Similarity 24.3%; Pred. No. 1,2e-10;
Matches 93; Conservative 90; Mismatches 148; Indels 52; Gaps 15;

```

```

QY 7 FPDFAFCRDTAHTGVEVLRLDGRKMKDVEELLRQRAQAEERYGKELVQIARAGG 66
Db 21 FGEHFWGE---KNHGFVLYHSVQGPSTKELADFTREATIEEYLSAMAKLSLAEN 77
QY 67 QTEMNSLRSPSLKQOTENVSAGHIQALALBELRLSLEEFEREROKERYE---AI 122
Db 78 GPMGTFAPLMEFVRVSDKLALCHLELTRKLDDILKRLRYGEEDLTKHKCKEEVST 137
QY 123 MDVQKSLSTLYKKTW-ESKKAVDQCRDADAEQAFEEVSNAGHOKVEKSONKAKOK 181
Db 138 LDVAV--VLSSGSQLPKRENTLNRCMD---QERLREESTS--QKEMDKMETYTKKKA 189
QY 182 ESATLERYRONIEOLERATEMEQEHRTTCEAFOLQFDRLLTLRNALWCHNOLSMQ 241
Db 190 ES-----LRSVSKYNSARADFEQKMLDSALRFQAMETHLIRMKALLGSAHSVET 242
QY 242 CVKDELVEEVLTLLEGCDVEGDINGFIOSKSTGEPAPVPYQNYVREYPLIGPSI 301
Db 243 HVQIGVHEEFKQNIENVSVELLKRFAESKGTGREKPGPLDFEAY-----SAAAL 293

```

```

QY 302 QPSCGVTKRFSGL----LHG-SPKTTSPAPASTETILP-----TPERNELVYASIEVQAT 352
Db 294 QEA---MKRLGAKAFRLPGLSREREPPEPAAVDFLEPDSGTCEVEDEGF--TVRDPVT 349
QY 353 QGNLNSSAODYRALLYDYTAONSD 375
Db 350 Q---NSTAEPSR-----FSSSDSD 365

```

```

RESULT 8
ID 0920W5 PRELIMINARY; PRT; 441 AA.
AC 0920W5;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE STNDAPIN I.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;
RX MEDLINE; 99135893.
RA QUALMANN B., ROOS J., DIGREGORIO P., KELLY R.B.;
RT "Synadin I, a synaptic dynamin-binding protein that associates with
RT the neural Wiskott-Aldrich syndrome protein."
RL MOL. Biol. Cell 10:501-513(1999).
DR EMBL; AF104402; AAD16887.1; -;
DR HSSP; Q60631; 4GB0.
SQ SEQUENCE 441 AA; 50449 MW; 2D77A1A8 CRC32;

```

```

Query Match 12.4% Score 266.5; DB 11; Length 441;
Best local Similarity 21.3%; Pred. No. 6,2e-10;
Matches 99; Conservative 82; Mismatches 202; Indels 81; Gaps 15;

```

```

QY 3 AQLQFDFAFCRDTAHTG-YEVLRLDGRKMKDVEELLRQRAQAEERYGKELVQIA 61
Db 8 ASPEITDSFW-----EVGNKYKTVKIRIDGHRCLNDLMSCVQERAKTIKAYAOQLTDM 61
QY 62 RK-----AGGTEMNSLRSPSLKQOTENVSAGHIQALALBELRLSLEEFEREROK 114
Db 62 KRWROLIEKGPQ--YGLSTRWANGAMTTEADKYSLEHGEKNSLNE--DLEKVMQMDA 117
QY 115 -----QRRKYEAINDRYOKSLSTLYKKTMSKRAYDQKCDADDA--EQAFERYS 162
Db 118 YHKQIMGKFKETKEADGGRKAKQPKWAKMKELAKKAYHLACKEEKLAMTRENNSKTE 177
QY 163 ANGHOKVEKSONKAKOKCESATEAERYRONIEOLERATEMEQEHRTTCEAFOLQFED 222
Db 178 QSVTFEQOKLVDKVDKQHDVQKQEKYKVLVDVGTTPPYMGSMQVPEQCOQFEK 237
QY 223 RLTLIR-----NALWYCHNOLSMQCYKDELVEEVLTLLEGCDVEGDINGF 268
Db 238 RLVFLEKVELLDIKRLHNLTAENSSYLH-----VYRELEQALRGADAQEDLRWF 284
QY 269 IQSKSTGREPAPVPYQNYVREYPLIGSPSIQPSGVTKRFSGLSGSKTTSPSAA 328
Db 285 RSTSPGM--PAINQOFEEWNPDLHTLAKEKQPK---KAEGALSNAGAVESTSOA 338
QY 329 STELTTPPERNELVYASIEVQATGN-----LNSSAODY-----RALYDYTAON 373
Db 339 GDRGSVSSYDRQO--AYATEWSDDESGNFGCNEANNGANPNPEDAKKGYRVALLDYDQ 397
QY 374 SDELISAGDILVILEGEGDWMTVER--NGQGFVPGSTLEKL 415
Db 398 QDELSTFKAGDELTKLGEDEDEQGWCKGRLDGOLGLYPANYVEAI 441

```

```

RESULT 9
ID 013154 PRELIMINARY; PRT; 448 AA.

```


AC 013154;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE P52.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauaria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 97435298.
RA MERILAINEN J., LEHTO V.P., WASENITUS V.M.;
RT "FAP52, a novel, SH3 domain-containing focal adhesion protein.";
RL J. Biol. Chem. 272:23278-23284(1997).
DR EMBL; 250798; CAA90678.1; -
DR HSSP; 060631; 4GBQ.
DR PFAM; PF00611; FCH; 1.
DR PFAM; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
SQ SEQUENCE 448 AA; 51971 MW; 5B3C6CHD CRC32;

Query Match 12.4%; Score 266.5; DB 13; Length 448;
Best Local Similarity 20.6%; Pred. No. 6.3e-10;
Matches 97; Conservative 82; Mismatches 189; Indels 103; Gaps 12;

QY 9 DAFMCDFAHTG-TEVLLQRLDGRKMKDVEELLRQRAQAEERYGKELVQIARKAGQ 67
DB 16 DDFW-----EYGNVRYKRYRIDGHRCLNDLMNCIHERRIEYVAQQLTEWAKRKQL 69
QY 68 TE-----MNSLRTSFDLSIKOOTENYVSAHIQALALREELSLSEFFREROKE- 114
DB 70 VERGPQYGVTERAMCAFMSAEKYSSELHLEVKSLMNE--DFEKIKWQKEAFHKQMG 127
QY 115 --QRKYEALIMRVOKSKLSLYKTKMESKRAYOKCRDADA--EQAFERYVSANGHOKV 170
DB 128 FETKTEAEQCFRKAQRPMAKKLKEVEBAKRAYHAACKEEKRLAISRETNSSADPALNEQL 187
QY 171 EKSQNAKCKCESATEAERYRONIEOLERAPTEMEBOEHRTTCEAPLOLEFDRITLRNA 230
DB 188 KLIQDVEKSKQVLTAKTKAKTEKSLKELDNATPOYEMENQOVECCQFEKRLRFRREV 247
QY 231 LNVHCNQLSMQCVKD-DELYEEVRLTEGCDVEGDIINGFIQSKSTGREPPAPVYONYD 289
DB 248 LLEVQKHLDSLNVASYKNYIRELEQNIKTADAVEDLRFMRANOGPKMS---MMWPFQED 303
QY 290 REVTLIGSPSTIOPSCGVIKRFGSLHGSPTTPSAPAASTETLTPPEKNELVYASIEV 349
DB 304 DE-----WSADLNRLTSLRREKKKASDGVTLTGINGTD----- 336
QY 350 QATQGNLSSAODY----- 366
DB 337 QYQAPKHSVSSYEKENQSPYDMSDESNPPSSITDAKDTNPFEDDTSFVMEVRYRAL 396
QY 367 YDYTAQNSDELDISAGDILAVILEGEDGMWTVYR--NGQRFVPGSYLEKL 415
DB 397 YDEEGQEDLSEFKAGDELTKMENEDQGCWKGRLDNGVGLPFRANYVEPI 447

RESULT 10
ID 061644 PRELIMINARY: PRT; 441 AA.
AC 061644;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE H74 PROTEIN.
GN H74.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;
RA PIOMANN M.R.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X85124; CAA59437.1; -
DR HSSP; Q60631; 4GBQ.
DR PFAM; PF00611; FCH; 1.
DR PFAM; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
SQ SEQUENCE 441 AA; 50575 MW; 44E61BOE CRC32;

Query Match 12.3%; Score 265.5; DB 11; Length 441;
Best Local Similarity 21.2%; Pred. No. 7.2e-10;
Matches 100; Conservative 84; Mismatches 192; Indels 95; Gaps 16;

QY 3 AQLOPDAFWCDFTAHNG-TEVLLQRLDGRKMKDVEELLRQRAQAEERYGKELVQIA 61
DB 8 ASEBITDSFW-----EYGNVRYKRYRIDGHRCLNDLMSCVOERAKTEKAYAOQLTDMWA 61
QY 62 RK-----AGGOTEMNSLRTSFDLSIKOOTENYVSAHIQALALREELSLSEFFREROKE- 114
DB 62 KRWROLIEKGFQ--YGSLEKAWGAMTEADYSELHDEVKKSLNE--DLKVKWKQDA 117
QY 115 -----QRKYEALIMRVOKSKLSLYKTKMESKRAYOKCRDADA--EQAFERYVS 162
DB 118 YHQIMGSEFKETKEADEGFRKAQRPMAKKKMLEAKKAYHILACKERBLATRENSKTE 177
QY 163 ANGHQKQVEKQNAKCKCESATEAERYRONIEOLERAPTEMEBOEHRTTCEAPLOLEFD 222
DB 178 QSVTPQOQKLVDRKQRODVOKTOEKYERKVLDEYVCTKTPQYMGEMQOVECCQFEK 237
QY 223 RLTLIR-----NALWVCNQLSMQCVKDELYEEVRLTEGCDVEGDIINGF 268
DB 238 RLVLFLKEVLIDIRKHLNLAENSSYTH-----VRELEQAIRGADQEDLWKF 284
QY 269 IQSKSTGREPPAPVYONYDREVTPLIGSPSTIOPSCGVIKRFGSLHGSPTTPSAPAA 328
DB 285 RSTSGPGK--PMNMPQFEWNPDLPHHTAKKEKP-----KKAEG-----ATLSNMTG 330
QY 329 STELTTPPENEL-----VVASIEVQATQGN-----LNSSAODY-----RAL 366
DB 331 AVESTSQAGDSVSSYDRGQTVATEWSDDESQNPFGNEANGANPEDDAKGVRYAL 390
QY 367 YDYTAQNSDELDISAGDILAVILEGEDGMWTVYR--NGQRFVPGSYLEKL 415
DB 391 YDYGQEDDELSEFKAGDELTKGEEDEQGCWKGRLDNSQGLYFRANYVEAI 441

RESULT 11
ID 060861 PRELIMINARY: PRT; 336 AA.
AC 060861;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE GAS7 PROTEIN.
GN GAS7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN (CEREBELLUM);
RA CHAO C.K., KUO J.T., JU Y.T., LIN-CHAO S.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ224876; CAA12177.1; -
DR PFAM; PF00611; FCH; 1.
SQ SEQUENCE 336 AA; 39191 MW; 6BC7E677 CRC32;

Query Match 11.1%; Score 238.5; DB 4; Length 336;
Best Local Similarity 24.6%; Pred. No. 2.8e-08;

RESULT	12		
043144			
ID	043144	PRELIMINARY;	PRT; 412 AA.
AC	043144;		
DT	01-JUN-1998	(TREMBLrel. 06, Created)	
DT	01-JUN-1998	(TREMBLrel. 06, Last sequence update)	
DT	01-NOV-1999	(TREMBLrel. 12, Last annotation update)	
DE	KIAA0394.		
GN	KIAA0394.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Cranista; Vertebrata; Mammalia;		
OC	Eutheria; Primates; Carnivora; Hominoidea; Homo.		
LN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE-BRAIN;		
RA	ISHIKAWA K., NAGASE T., NAKAJIMA D., SEKI N., OHIRA M., MIYAJIMA N.,		
RA	TANAKA A., KOTANI H., NOMURA N., OHARA O.,		
LN	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AB007854; BAA23690.1; -		
DR	HSSP: Q13526; IPIN.		
DR	PFAM: PF00611; FCH: 1.		
DR	PFAM: PF00397; WW_rsp5_WWP; 1.		
DR	PRINTS: PR00403; WMDOMAIN.		
QO	SEQUENCE 412 AA; 47266 MW; 537604A2 CRC32;		

[illegible][illegible]

RESULT 14		
ID	055148	
AC	055148	PRELIMINARY;
DT	01-JUN-1998	(TREMBlrel. 06, Created)
DT	01-JUN-1998	(TREMBlrel. 06, Last sequence update)
DT	01-NOV-1999	(TREMBlrel. 12, Last annotation update)
DE	GAS-7 PROTEIN.	
GN	GAS-7.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;	
RA	CHAO C.C., LIN-CHAO S.;	
RL	Submitted (DEC-1997) to the EMBL/Genbank/DBD databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=CEREBELLUM;	
RA	CHAO C.C., KUO J.D., SU L.-J., LIU C.H., LIH C.-J., COHEN S.N.,	

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2000, 10:03:46 ; Search time 12.17 Seconds

(Without alignments)
807.702 Million cell updates/sec

Title: US-09-068-377-1

Sequence: 1 MMAQLQFRDAFWCRDFAHT.....WTVNRNGRGVPGSYLEKL 415

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2152	100.0	415	W71595	Murine tyrosine ph
2	265.5	12.3	441	W05410	Mouse H74 protein.
3	197	9.2	377	W05411	Human H74 protein.
4	181	8.4	788	W05393	Mouse SH3P12 prote
5	179.5	8.3	433	W05389	Mouse SH3P7 protei
6	170.5	7.9	434	W05391	Mouse SH3P9 protei
7	167	7.8	355	W05406	Mouse cortactin pr
8	153	7.1	355	W05398	Human clone 56 pro
9	150.5	7.0	451	W05602	Human box-dependen
10	149.5	6.9	451	W47295	Human Bin1. Box-de
11	149.5	6.9	451	W45004	Human Bin1. Box-de
12	149	6.9	1484	W89721	Canine ribosome re
13	148.5	6.9	509	W05399	Human clone 65 pro
14	148.5	6.9	534	W82799	Human IRS-P53h pro
15	147.5	6.9	2101	R47173	Sequence of the in
16	145.5	6.8	182	W05402	Human clone 53 pro
17	144	6.7	205	W05407	Mouse M1N50 protei
18	144	6.7	261	W25767	Human lasp-1. Gene
19	143	6.6	929	W14593	Streptococcus pneu
20	142.5	6.6	2192	W21732	LexA/NuA fusion p
21	142.5	6.6	2272	W21731	GAL4/HuA/NuA fusio
22	139.5	6.5	404	W05392	Human SH3P9 protei
23	137	6.4	359	W37962	Amino acid sequenc
24	137	6.4	553	W25115	CD2-associated int
25	137	6.4	553	W25495	CD2-associated int
26	137	6.4	553	W80419	CD2-associated int
27	137	6.4	659	W37724	CD2-associated int
28	136	6.3	324	W25117	CD2-associated int
29	136	6.3	324	W26497	CD2-associated int
30	136	6.3	324	W80421	CD2-associated int
31	134.5	6.2	390	R83825	p47(phox) protein.
32	134	6.2	486	Y03795	Murine Lck binding
33	134	6.2	1392	Y06999	Restin protein seq

34	133	6.2	462	1	R05766	Portion of peptide
35	133	6.2	1427	1	R10534	Human 160KD mediat
36	132	6.1	619	1	R63437	Pneumococcal surfa
37	132	6.1	619	1	R86911	Pneumococcal surfa
38	132	6.1	619	1	R87598	Pneumococcal surfa
39	132	6.1	648	1	W62274	Streptococcus pneu
40	132	6.1	648	1	W70336	Pneumococcal surfa
41	132	6.1	648	1	W87879	A pneumococcal sur
42	132	6.1	653	1	W92456	S. pneumoniae trun
43	132	6.1	684	1	R73912	Streptococcus pneu
44	130.5	6.1	464	1	W25116	CD2-associated int
45	130.5	6.1	464	1	W26496	CD2-associated int

ALIGNMENTS

RESULT	ID	Score	Query Match	Length	ID	Description
1	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
2	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
3	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
4	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
5	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
6	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
7	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
8	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
9	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
10	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
11	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
12	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
13	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
14	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
15	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
16	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
17	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
18	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
19	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
20	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
21	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
22	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
23	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
24	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
25	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
26	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
27	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
28	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
29	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
30	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
31	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
32	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
33	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
34	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
35	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
36	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
37	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
38	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
39	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
40	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
41	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
42	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
43	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
44	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -
45	W71595	100.0%	Score 2152; DB 1; Best local similarity 100.0%; Pred. No. 6e-185; Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	415	W71595	Murine tyrosine phosphatase interacting polypeptide -

```

Db 241 QCVKDELVEEKLLEGCDVEEDINGFIQSKSTGREPPAPVQYKYREYETPLIGSS 300
QY 301 IOPSCGVIRFSGLIHGSPKTPSPAPASTETLTPPENNELVYASIEVQAOGLNNSA 360
Db 301 IOPSCGVIRFSGLIHGSPKTPSPAPASTETLTPPENNELVYASIEVQAOGLNNSA 360
QY 361 QDYRALDYTAONSDDELISAGDILAVILEGEDGWTVERNGORGFVPGSYLEKL 415
Db 361 QDYRALDYTAONSDDELISAGDILAVILEGEDGWTVERNGORGFVPGSYLEKL 415

RESULT 2
W05410
ID W05410 standard; protein; 441 AA.
AC W05410;
DT 23-FEB-1998 (first entry)
DE Mouse H74 protein.
KW Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
cellular signalling element; cellular structural element; malignancy;
protein identification; functional domain; protein screening;
cellular signal transduction process.
OS Mus musculus.
FH Key Location/Qualifiers
FT Misc_difference 439 /note= "encoded by GAC"
FM W09631625-A1.
PD 10-OCT-1996.
PE 04-APR-1996; U04454.
PR 03-APR-1996; US-630915.
PR 07-APR-1995; US-417872.
PA (CYTO-) CYTOGEN CORP.
PI (UYN-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;
DR N-PSDB; T39809.
PT Identifying polypeptide(s) having specific functional domain (esp.
PT SH3 domain) - comprises detecting selective binding to recognition
PT unit, regardless of sequence homology
PS Claim 102; Fig 47; 174pp; English.
CC W05405-W05411 represent human and mouse Src-homology region 3 (SH3)
CC domain containing proteins that can be used in the method of the
CC invention. SH3 domain containing proteins play a role in signalling and
CC structural elements of cells. The method of the invention is for
CC identifying polypeptides containing functional domains of interest
CC (especially SH3 domains). The method comprises contacting a multivalent
CC recognition unit (RU) complex with a number of peptides and identifying
CC polypeptides having a selective binding affinity for the RU complex. The
CC method is based on functional similarities and does not rely on sequence
CC similarities. Prior methods only gave limited success for identifying
CC proteins which contain an SH3 domain due to the minimal sequence homology
CC among known SH3 proteins. It has been found that small peptide RUs in
CC multivalent form have reduced specificity for a given functional domain
CC compared to monomer RUs. Multivalent RU complexes are particularly suited
CC to screening for polypeptides containing functional domains that are
CC similar to, but not identical in sequence to, the original target
CC functional domain. The new method enables proteins having a common
CC function to be identified. Identification of novel SH3 proteins will be
CC useful for a better understanding of cell growth, malignancy, signal
CC transduction processes, etc. New candidate drugs can be identified, and
CC their specificities (e.g. pharmacological activities) can be assessed
CC using the method of the invention.
SQ Sequence 441 AA;

```

```

Query Match 12.3%; Score 265.5; DB 1; Length 441;
Best Local Similarity 21.2%; Pred. No. 1.1e-15;
Matches 100; Conservative 84; Mismatches 192; Indels 95; Gaps 16;

```

```

QY 3 AQLQRFDAFWCRDFTAHNG-YEVLLDRLDGKMKCKDVBEILLRQAOAEERKGLVQIA 61
DB 8 ASEETIDSEFW-----EVGNKRYKRYKRIDDGHRRLCDNLMSGVQERAKLEKAYADLTQWA 61

```

```

QY 62 RK-----AGQTEENSLRTSFDLSKQGTENVGSAHIQALALREELRSLSEFFERQKE- 114
Db 62 KRWROLIEKGPQ--YGSIERAMGAMMTADKVSLEHGVKNSLLNF--DLEKYNMOKDA 117
QY 115 -----QRKRYEALIMDVQKSKLSLYKKYTESKAYDQCKRADDA--EQAFERVS 162
Db 118 YHKQIMGFKEETKEKEDGFERRAKQPMKKMKELEPAKRAVLAHCKEERLMTREMSKTE 177
QY 163 ANGHQKQVEKSONKAKQCKESATAEERYRONIEQLERARTEMBOEHRTCEAFQIQEFD 222
Db 178 QSVTPQOQKRLVDKVCRODVQKQEKYEVLVEDVCKTTPQYHEGMEQYFEOCQOFEK 237
QY 223 RLTLIR-----NALWVHCNQLSMQCKVDELVEEYRLTLEGCDVEEDINGF 268
Db 238 RLVLFLKEVLIDIKRHLMLAENSYSYH-----YRELEQAIIRADAOQEDLRWF 284
QY 269 IQSKSTGREPPAPVQYQNYDREYTPILIGSPSTQPSGVAKRSGLIHGSPKTPSPAPAA 328
Db 285 RSTGPGM--PMNMPQPEEMNPDPHTTAKKERQ-----KKAEG-----ATLSNATG 330
QY 329 STELTTPPERNEL-----VYASIEVQAOGLN-----LNSAODY-----RAL 366
Db 331 AVESTSQAGDRGSVSYDRGQYATEMSDESGNPFPGNANGAMPFEDDAKGVAVRAL 390
QY 367 YDYTAONSDDELISAGDILAVILEGEDGWTVER--NGORGFVPGSYLEKL 415
Db 391 YDYDQGODELSEKAGDELTKLGEEDGQRCRGLDGLGYPANYVAI 441

RESULT 3
W05411
ID W05411 standard; protein; 377 AA.
AC W05411;
DT 23-FEB-1998 (first entry)
DE Human H74 protein.
KW Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
cellular signalling element; cellular structural element; malignancy;
protein identification; functional domain; protein screening;
cellular signal transduction process.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 1 /note= "encoded by GA"
FM W09631625-A1.
PD 10-OCT-1996.
PE 04-APR-1996; U04454.
PR 03-APR-1996; US-630915.
PR 07-APR-1995; US-417872.
PA (CYTO-) CYTOGEN CORP.
PI (UYN-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;
DR N-PSDB; T39810.
PT Identifying polypeptide(s) having specific functional domain (esp.
PT SH3 domain) - comprises detecting selective binding to recognition
PT unit, regardless of sequence homology
PS Claim 102; Fig 49; 174pp; English.
CC W05405-W05411 represent human and mouse Src-homology region 3 (SH3)
CC domain containing proteins that can be used in the method of the
CC invention. SH3 domain containing proteins play a role in signalling and
CC structural elements of cells. The method of the invention is for
CC identifying polypeptides containing functional domains of interest
CC (especially SH3 domains). The method comprises contacting a multivalent
CC recognition unit (RU) complex with a number of peptides and identifying
CC polypeptides having a selective binding affinity for the RU complex. The
CC method is based on functional similarities and does not rely on sequence
CC similarities. Prior methods only gave limited success for identifying
CC proteins which contain an SH3 domain due to the minimal sequence homology
CC among known SH3 proteins. It has been found that small peptide RUs in
CC multivalent form have reduced specificity for a given functional domain
CC compared to monomer RUs. Multivalent RU complexes are particularly suited
CC to screening for polypeptides containing functional domains that are
CC similar to, but not identical in sequence to, the original target

```


PA (UYNC-) UNITV NORTH CAROLINA.
 PI Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;
 DR WPI; 96-465045/46.
 DR N-PSDB; T39789.
 PT Identifying polypeptide(s) having specific functional domain (esp.
 SH3 domain) - comprises detecting selective binding to recognition
 PT unit, regardless of sequence homology
 PS Claim 54; Fig 31; 174pp; English.
 CC W05386-W05403 represent novel human and mouse Src-homology region 3 (SH3)
 CC domain containing proteins that can be used in the method of the
 CC invention. SH3 domain containing proteins play a role in signalling and
 CC structural elements of cells. The method of the invention is for
 CC identifying polypeptides containing functional domains of interest
 CC (especially SH3 domains). The method comprises contacting a multivalent
 CC recognition unit (RU) complex with a number of peptides and identifying
 CC polypeptides having a selective binding affinity for the RU complex. The
 CC method is based on functional similarities and does not rely on sequence
 CC similarities. Prior methods only gave limited success for identifying
 CC proteins which contain an SH3 domain due to the minimal sequence homology
 CC among known SH3 proteins. It has been found that small peptide RUs in
 CC multivalent form have reduced specificity for a given functional domain
 CC compared to monomer RUs. Multivalent RU complexes are particularly suited
 CC to screening for polypeptides containing functional domains that are
 CC similar to, but not identical in sequence to, the original target
 CC functional domain. The new method enables proteins having a common
 CC function to be identified. Identification of novel SH3 proteins will be
 CC useful for a better understanding of cell growth, malignancy, signal
 CC transduction processes, etc. New candidate drugs can be identified, and
 CC their specificities (e.g. pharmacological activities) can be assessed
 CC using the method of the invention.
 SO Sequence 433 AA;

Query Match 8.3%; Score 179.5; DB 1; Length 433;
 Best Local Similarity 20.2%; Pred. No. 5.4e-08;
 Matches 80; Conservative 59; Mismatches 145; Indels 113; Gaps 14;

OY 47 AOAERYGKELV--QIARKAGGOTEMSLTSPDSLKQOTENVGSAHTQLALALREELRS 104
 DB 117 ARAEDEVPEPCIMEKAKAGANYSFKESTSFODVQPAQ-VGSYV-----QKTNA 167
 OY 105 LEEFEROKORRKYKKAIMRVOKSKLSLYKKTWESKKAVDOKCRADDAEQAPEVSAN 164
 DB 168 ISEIK-----RVCKDNFMKAKEKEENRRLIEKRR-----AEEERORLIEE 208
 OY 165 GHOKVOKESONKAKOKE---SATEAEVRYRONIEOLERARTEWEOHRTTCEAFOLSE 221
 DB 209 RRERELQEAARRRQYOEHRISAGAPSRTEPEQEAVSRTROWE----- 253
 OY 222 DRLLTRNALMWHGNOLOSMOCVKKDELVEVRLLTEGCOVEGDIING---PIQSKST-- 274
 DB 254 -----SAGQDAHPREIFRFOKERAAMSTTVTSQPEKLSPLQKLTGP 298
 OY 275 -----GEPAPVPYQNYVREVTPLIGSPISQGVIRFSGILGSPKTPPSAPAAST 330
 DB 299 EMTSGREPTAPYS-----PPAGVCE-----EPARSTISSAATESE 334
 OY 331 ETLTPPENNELVYASIEVQATOGNINSSAOY-----RALITYTAONSDLE 377
 DB 335 PLYEVPEEDTL-YEPPPLVQOQAGSEHIDNWMOSQSGOGLCARALYDYAADDEI 393
 OY 378 DISAGDILAVILEGEDGWT-VERNGORGFVPGSYLE 413
 DB 394 SPDPENLITGIEVIDEGMWGRYGGPGHPEMPFANYVE 430

RESULT 6
 W05391
 ID W05391 standard; Protein; 434 AA.
 AC W05391;
 DT 18-FEB-1998 (first entry)
 DE Mouse SH3p9 protein.
 KW Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;

KW cellular signalling element; cellular structural element; malignancy;
 KW protein identification; functional domain; protein screening;
 KW cellular signal transduction process.
 OS Mus musculus.
 FH Key Location/Qualifiers
 FT Misc-difference 433
 PN W09631625-A1.
 PD 10-OCT-1996.
 PE 04-APR-1996; U04454.
 PR 03-APR-1996; US-630915.
 PR 07-APR-1995; US-417872.
 PA (CYTO-) CYTOGEN CORP.
 PA (UYNC-) UNITV NORTH CAROLINA.
 PI Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;
 DR WPI; 96-465045/46.
 DR N-PSDB; T39791.
 PT Identifying polypeptide(s) having specific functional domain (esp.
 SH3 domain) - comprises detecting selective binding to recognition
 PT unit, regardless of sequence homology
 PS Claim 54; Fig 35; 174pp; English.
 CC W05386-W05403 represent novel human and mouse Src-homology region 3 (SH3)
 CC domain containing proteins that can be used in the method of the
 CC invention. SH3 domain containing proteins play a role in signalling and
 CC structural elements of cells. The method of the invention is for
 CC identifying polypeptides containing functional domains of interest
 CC (especially SH3 domains). The method comprises contacting a multivalent
 CC recognition unit (RU) complex with a number of peptides and identifying
 CC polypeptides having a selective binding affinity for the RU complex. The
 CC method is based on functional similarities and does not rely on sequence
 CC similarities. Prior methods only gave limited success for identifying
 CC proteins which contain an SH3 domain due to the minimal sequence homology
 CC among known SH3 proteins. It has been found that small peptide RUs in
 CC multivalent form have reduced specificity for a given functional domain
 CC compared to monomer RUs. Multivalent RU complexes are particularly suited
 CC to screening for polypeptides containing functional domains that are
 CC similar to, but not identical in sequence to, the original target
 CC functional domain. The new method enables proteins having a common
 CC function to be identified. Identification of novel SH3 proteins will be
 CC useful for a better understanding of cell growth, malignancy, signal
 CC transduction processes, etc. New candidate drugs can be identified, and
 CC their specificities (e.g. pharmacological activities) can be assessed
 CC using the method of the invention.
 SO Sequence 434 AA;

Query Match 7.9%; Score 170.5; DB 1; Length 434;
 Best Local Similarity 21.0%; Pred. No. 3.5e-07;
 Matches 96; Conservative 73; Mismatches 171; Indels 117; Gaps 19;

OY 33 RAKCDVVELLQORQAERYGKELVQIARKAGGQ-TE---MNSIRSFPSLKQOTENV 87
 DB 20 KILTRAQEKVLOKLGKADPTEKDEQECQVQNFKNQLOLGTGLQDLKRYLASVAMHE-- 77
 OY 88 GSAHTQLALALREELRSLEEEREROKRKYEA-----MDRVOKSKLSLYKKTME 139
 DB 78 -----ASKKLSECLQEVYEPFEMRDEANKIAENNDLMDYDQK-----LVDDALL 124
 OY 140 SKKAY-----DQCRDDDAEQAPEFVSANGHQOYE---KSONKAKOKESTAEVRY 190
 DB 125 TMDYTLGQFPPIKSIARKGRKLLVDYSARHHYESTLQYAKKKDAKLAKEEELIKAKOY 184
 OY 191 YRQ-NIEOLERARTEWEOHRTTCEAFOL-----QEFDRLLTRNALMWH----- 234
 DB 185 FEEMNVDLQELPGLMSRVGFYVNSFQSLAGLEENHKKESKLNQNLNDVLSLEKQNG 244
 OY 235 CNOLSMOCVKKDELVE-----EVRLLTEGCDVEGDINGFIQOSKSTGR 276
 DB 245 SNTFTYKAQPSDNPAREKGNKSPSPDPGSPAATPEIRVNH-----PEPASGA 292
 OY 277 EPPAPVPYQNYVREVTPLIGSPISQGVIRFSGILGSPKTPPSAPAASTETLTP 336
 DB 293 SPGATITKSPSPQAEASEVVGAAQ-EP-----GETAASATSSSLPAVVVFESAT 342


```

OY 337 PERNELVASEIVQATQGNLTN--SSAQDYRALYDTAANCSELDASDIIALVI----- 388
DB 343 -----VNCAGVSGSAGTGDLDPDPGFMFVVOAQHDYATDITDELQAKADVYLVYIFQNP 396
OY 389 LEGEDGW-----WTVERNQG--RGVPVGSYLEKL 415
DB 397 ERODEGWLIMGVKESDMNQHEKLEKRGVPEPNTFERY 433
RESULT 7
ID W05406 standard; Protein: 546 AA.
AC W05406;
DT 20-FEB-1998 (first entry)
DE Mouse coractin protein.
KW Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
KW cellular signalling element; cellular structural element; malignancy;
KW protein identification; functional domain; protein screening;
KW cellular signal transduction process.
OS Mus musculus.
PE Key Location/Qualifiers
FT Misc_difference 385
FT /note= "encoded by CAG"
PN W0631625-A1.
PD 10-OCT-1996.
PD 04-APR-1996; U04454.
PR 03-APR-1996; US-630915.
PR 07-APR-1995; US-417872.
PA (CYTO-) CYTOGEN CORP.
PA (DYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DK, Hoffman N, Kay BK, McConnell SJ, Sparks AB;
PI WPI: 96-465045/46.
DR N-PSDB: 7139805.
PT Identifying polypeptide(s) having specific functional domain (esp.
PT SH3 domain) - comprises detecting selective binding to recognition
PT unit, regardless of sequence homology
PS Claim 102; Fig 27; 174pp; English.
SC W05405-W05411 represent human and mouse Src-homology region 3 (SH3)
CC domain containing proteins that can be used in the method of the
CC invention. SH3 domain containing proteins play a role in signalling and
CC structural elements of cells. The method of the invention is for
CC identifying polypeptides containing functional domains of interest
CC (especially SH3 domains). The method comprises contacting a multivalent
CC recognition unit (RU) complex with a number of peptides and identifying
CC polypeptides having a selective binding affinity for the RU complex. The
CC method is based on functional similarities and does not rely on sequence
CC similarities. Prior methods only gave limited success for identifying
CC proteins which contain an SH3 domain due to the minimal sequence homology
CC among known SH3 proteins. It has been found that small peptide RUs in
CC multivalent form have reduced specificity for a given functional domain
CC compared to monomer RUs. Multivalent RU complexes are particularly suited
CC to screening for polypeptides containing functional domains that are
CC similar to, but not identical in sequence to, the original target
CC functional domain. The new method enables proteins having a common
CC function to be identified. Identification of novel SH3 proteins will be
CC useful for a better understanding of cell growth, malignancy, signal
CC transduction processes, etc. New candidate drugs can be identified, and
CC their specificities (e.g. pharmacological activities) can be assessed
CC using the method of the invention.
SQ Sequence 546 AA;
Query Match 7.8%, Score 167, DB 1; Length 546;
Best Local Similarity 23.2%, Fred. No. 1e-06;
Matches 92; Conservative 41; Mismatches 135; Indels 128; Gaps 15;
OY 40 EELLKQRAQAEERYKELVQIRKAGCGGTENMSLRSPFSLQGTENWVSAHIQALALR 99
DB 254 EKLQIHESQKQKVTG-----FGCKRGVQSE-----KQDSANVGFDV-----K 290
OY 100 EELNSLEEFERQKQKRYEALIMRVQKSLTYLKTKTESKRAYDQKCRADDAEQAPE 159

```

```

Db      291 ERLAKKEPQODYAKGFGKGYGVQKDRMDKN-ASTFEEVGVPSAY----- 334
QY      160 RVSANGHQVQVEKSOAKAKQCKESATAEAEVYQNIQLEPRAETWEQEHRTTCPEAFOLO 219
Db      335 -----QKTVP-----IEAVTSKTSNIANFENLAKER--EODERRKAEEARRAQ 375
QY      220 EFDRLITLNNALWVHCNQLSMQCVKQDELYEEVRLTEGCDVEDDINGFIQSKSTGREPP 279
Db      376 -----RMAKERQDEBEARKLIE-----QARRKQTPP 403
QY      280 AEPVYQNIYDR-----EVPPLICSPSIQSPCGVIKRPSGLLHSGPKTPPSADA-- 328
Db      404 ASPSPQPIDRPPSPSPTEYDAAPFAKPEPSYRGS-----EPEEYSIEAGAP 450
QY      329 ---STPELTPEERNLVLVYASIEV---QATQGNINSSADYR---ALVDYTAQNSDEL 377
Db      451 EAGSQOGLTYT---SEPVETETAPRPHYQAEDDTTYDESDLGITATALDYQAAGDEI 507
QY      378 DISAGDILAVILEGEDGMWTVERNNGRGFPYGSYLE 413
Db      508 SFDPDITINIEIMIDGWMRGVCKGRYGLPEPAYVE 543

RESULT      8
ID      W05398
AC      W05398:
DT      19-FEB-1998 (first entry)
DE      Human clone 56 protein.
KW      Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
KW      cellular signalling element; cellular structural element; malignancy;
KW      protein identification; functional domain; protein screening;
KW      cellular signal transduction process.
OS      Homo sapiens.
PN      MO9631625-A1.
PD      10-OCT-1996.
PF      04-APR-1996; U04454.
PR      03-APR-1996; US-630915.
PR      07-APR-1995; US-417872.
PA      (CYTO-) CYTOGEN CORP.
PA      (UYNC-) UNIT NORTH CAROLINA.
PI      Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB:
DR      WPI: 96-465045/46.
DR      N-PDB: T39796.
PT      Identifying polypeptide(s) having specific functional domain (esp.
PT      SH3 domain) - comprises detecting selective binding to recognition
PT      unit, regardless of sequence homology
PS      Claim 54: Fig 57: 17app: English.
CC      W05386-W05403 represent novel human and mouse Src-homology region 3 (SH3)
CC      domain containing proteins that can be used in the method of the
CC      invention. SH3 domain containing proteins play a role in signalling and
CC      structural elements of cells. The method of the invention is for
CC      identifying polypeptides containing functional domains of interest
CC      (especially SH3 domains). The method comprises contacting a multivalent
CC      recognition unit (RU) complex with a number of peptides and identifying
CC      polypeptides having a selective binding affinity for the RU complex. The
CC      method is based on functional similarities and does not rely on sequence
CC      similarities. Prior methods only gave limited success for identifying
CC      proteins which contain an SH3 domain due to the minimal sequence homology
CC      among known SH3 proteins. It has been found that small peptide RUs in
CC      multivalent form have reduced specificity for a given functional domain
CC      compared to monomer RUs. Multivalent RU complexes are particularly suited
CC      to screening for polypeptides containing functional domains that are
CC      similar to, but not identical in sequence to, the original target
CC      functional domain. The new method enables proteins having a common
CC      function to be identified. Identification of novel SH3 proteins will be
CC      useful for a better understanding of cell growth, malignancy, signal
CC      transduction processes, etc. New candidate drugs can be identified, and
CC      their specificities (e.g. pharmacological activities) can be assessed
CC      using the method of the invention.
Sequence      355 Aa;

```

Query Match 7.1%; Score 153; DB 1; Length 355;
 Best Local Similarity 20.5%; Pred. No. 9.5e-06;
 Matches 60; Conservative 42; Mismatches 120; Indels 70; Gaps 9;

QY 147 KCRDADDAEQAERVSANGHOKVKSQKAKCKESATEAREYRONILEPATERME 206
 DB 45 KAKREKRNRLLEKMAEADQROLEQ-ERRERELREARBRORFOGEGASPOSRT-WE 102
 QY 207 QEHRTCEAFOLQEPDRLLILNALMWCHNLSMQCVKDELVEFVALLTEGCDVEDIN 266
 DB 103 QOOEYVSHNRNEQ-----SAVHPREIRKOKERASTSTSISSPOP 142
 QY 267 G-----FIQSKST-----GHEPPAPVYONYDREVTPLIGSPSIQPSGVIRKRSGL 315
 DB 143 GKLSAPFLQKOLQIPEFHGEPAAAIIS-----RPRADLPREP----- 181
 QY 316 HGSPTTPSPAASTETLTPTPERNELVYASIEVQATG-----NLNSAAD 362
 DB 182 --ASTPPCLVQAEAEVVEEPPEQETFYHQPPLVQOQAGSEHIDHNIQOGGLSGGLC 239
 QY 363 YRALYDTAONSDDELISAGDILAVILEGEGDGMW-VERNGRGFVPGSYLE 413
 DB 240 ARALYDQADDTETISFDPENLITGIEVIDEGWNGYGPDGHGMPAPAYVE 291

RESULT 9
 W06602 ID W06602 standard; Protein: 451 AA.
 AC W06602;
 DT 07-FEB-1997 (first entry)
 DE Human box-dependent myc-interacting protein (BIN1).
 KW Box-dependent myc-interacting protein; BIN1; MIP-99; oncogene;
 KW oncoprotein; breast cancer; liver cancer; apoptosis;
 OS tumour suppressor protein; hyperplasia; diagnosis; therapy.
 FH Key
 FT domain Location/qualifiers
 FT 1..222
 FT /note="polypeptide useful in constructs of the
 FT invention (Claim 9)"
 FT 190..250
 FT /note="polypeptide useful in constructs of the
 FT invention (Claim 9)"
 FT 223..251
 FT /note="polypeptide useful in constructs of the
 FT invention (Claim 9)"
 FT 252..265
 FT /label="Nuclear localisation domain
 FT /note="polypeptide useful in constructs of the
 FT invention (Claim 9)"
 FT 263..397
 FT /note="polypeptide useful in constructs of the
 FT invention (Claim 9)"
 FT 270..383
 FT /label="Myc-binding domain
 FT /note="polypeptide useful in constructs of the
 FT invention (Claim 9)"
 FT 278..451
 FT /label="SH3
 FT /note="SH3 domain useful in constructs of the
 FT invention (Claim 9); this domain may
 FT be useful in blocking apoptosis"
 FT domain
 FT domain
 PD W09634627-A1.
 PF 07-NOV-1996.
 PF 02-MAY-1996; U06231.
 PF 05-MAY-1995; US-435454.
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
 PI Prendergast GC, Sakamuro D;
 DR MPI; 96-505902/50.
 DR N-PSDB; T43573.
 PT Mammalian box-dependent myc-interacting protein Bin1 coding
 PT sequences - used in the diagnosis and treatment of c-myc associated
 PT cancers and hyperplastic disease and Bin1 deficiency
 PS Claim 3; Page 44-46; 79pp; English.

CC Human box-dependent myc-interacting protein Bin1 (W06602) is a
 CC putative tumour suppressor protein. It inhibits Myc-dependent
 CC malignant cell transformation, and expression of Bin1 mRNA is
 CC lost in human liver and breast carcinoma cell lines. The amino
 CC acid sequence of human Bin1 was deduced from a cDNA clone (T43573)
 CC isolated from a human skeletal muscle cDNA library. Recombinant
 CC Bin1 or its polypeptides can be produced in host cells for use in
 CC the detection, diagnosis and treatment of Bin1 deficiency for use in
 CC and cancers or hyperplastic diseases involving the c-myc oncogene.
 SQ Sequence 451 AA;

Query Match 7.0%; Score 150.5; DB 1; Length 451;
 Best Local Similarity 21.7%; Pred. No. 2.3e-05;
 Matches 100; Conservative 70; Mismatches 186; Indels 105; Gaps 20;

QY 33 KMKCKDVELLRQARQAEERYGKELVQIARKAGQ-TE---MNSLRFSLSKQOTENV 87
 DB 17 KKLTRAQEKVLOKIKGKADFTKDEQFQCVQNFNKLTEGTRLOKRLTYLASVKAMHE-- 74
 QY 88 GSAHIQALALREELRLSEEREROKKRYEAT-----MDRVQSKLSLYKKTME 139
 DB 75 -----ASKRLNECLQEVTEPDMPPGRDEANKIAENNDLMDYHOK---LVDQALL 121
 QY 140 SKKAY-----DQCRDADDAEQAERVSANGHOKVE---KSONKAKCKESATEAREY 190
 DB 122 TMDTYLGQFPDLSRIYAKGKRLVDYDSARHHYESLQAKKDEKAKAEELIKQKV 181
 QY 191 YRO-NIEOLERATEWEQHRITCEAFQ-----QEDRLTLNALMWCHNQ--- 237
 DB 182 FEEMVNDLQELPRLMNSVGVYVTFPOSIALEENFKREKSNQNDVVLGLERQHG 241
 QY 238 -----LSMQCVKDELVEEVRLLTEGCDVEDINGFIQSKSTGPREPPAPVYONYDREV 292
 DB 242 SNTFTYKQAPRRKSKLFSRLRKKNSDNAKGN---KSPSPGSGSAAP-ETRVNHEP 297
 QY 293 TPDLG-----SPS-----IQSCGVIRFSGLLGSKTTPSPAASTET 332
 DB 298 EPAGATPGATLPKSPSPAPAEAYAGSTOPAGA--QEGETASASASASPAAVYET 355
 QY 333 LTPPERNELVYASIEVQATGNNL---SSAODYRALYDTAONSDDELISAGDILAVI- 388
 DB 356 PPAT-----VNGTVEGSGAGRLDPPGFEMFKVQAOHDYATDTDLQKADVVAVLP 409
 QY 389 ---LEGEDGW-----WTERNGQ--RGYVPGSTLEKL 415
 DB 410 FQNPEDQEGHMGVKSQDMNQHKLEKRCRGVPEPNTTERV 450

RESULT 10
 W47295 ID W47295 standard; Protein: 451 AA.
 AC W47295;
 DT 01-JUN-1998 (first entry)
 DE Human Bin1.
 KW Human.
 KW Human; Box-dependent myc-interacting protein; Bin1;
 KW treatment; cancer; hyperplastic disease; neurodegenerative disease;
 KW antibody; diagnosis.
 OS Homo sapiens.
 PN US5723581-A.
 PD 03-MAR-1998.
 PF 24-MAY-1996; 652972.
 PF 24-MAY-1996; US-652972.
 PF 05-MAY-1995; US-435454.
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
 PI Prendergast GC, Sakamuro D;
 DR MPI; 98-178541/16.
 DR N-PSDB; V15692.
 PT Box-dependent myc-interacting protein Bin1 - useful for, e.g.
 PT treating diseases associated with inadequate Bin1 levels such as
 PT colon cancer
 PS Claim 1; Columns 25-30; 32pp; English.
 CC The present sequence is human Box-dependent

CC myc-interacting protein Bin1, which can be used to treat diseases
CC associated with inadequate Bin1 levels, e.g. liver, colon or breast
CC cancer or hyperplastic diseases, or diseases where Bin1 activity
CC needs to be inhibited, e.g. neurodegenerative diseases, or to raise
CC antibodies for diagnostic purposes.
CC Sequence 451 AA:

Query Match	6.98;	Score 149.5;	DB 1;	Length 451;
Best Local Similarity	21.78;	Pred. No. 2.8e-05;		
Matches 100; Conservative	70;	Mismatches 186;	Indels 105;	Gaps 20;

QY	33	KKMOCKDVBEILLROQAQAEERKGRGLVQIARAGQ-TE-----	ANSLRSTFSDLSLKQOTENY	87
Db	17	KLLPRAEQEYLOTKGKADDETKDEOFEECVONFNKOLTEGRLOKDTLYTASVAKANE--	74	
QY	88	GSAHQIALALREELRSLEEFEREROKORRKYEA1-----	MDRQOKSKLSLYUKTME	139
Db	75	-----ASKKINBELOQEYVEYBPDMWGRDBEANKIAENNDLIMMDYOK-----	LYVQALL	121
QY	140	SKKAY-----DQCRDADDEQOAEERYVSANGHOKOVE----	KSONKAKOCKESATEAERY	190
Db	122	TMDYTLQOPFDIKSRIRKGRKELVDYVSANHHNESLOTAKKBEAKTAKAEELIYAKQV	181	
QY	191	YRO-NIEOLEBARTQEMQEHRTTCEAROL-----	OEDRLTILRNALMYHCNO--	237
Db	182	FEENAVNDLOEBELSLMNSRQGVYVNTFOSTAGLEENHKKMSKINOMLNVGLKEKONG	241	
QY	238	-----LSMOCVKDELEYEVAVRLTLEGDCVBDINGFLOSKSTGSEPPAVYONYUDREV	292	
Db	242	SNTEYTKAOPRRKSKSLRSRLRRKKNSDNAPAKGN-----	KSPSPDGPSPATR-ELRVNHER	297
QY	293	TPLLG-----SPS-----	IQPSGCYIKRFSGLHGSKTTTPSPAASSTET	332
Db	298	EPAGGATPGATLPKPSPOPADESVAGCTPAGA--OEBGETPAASAASSLSIPAVVET	355	
QY	333	LTPPERNEULYASIEVOATQGNLN-----SSAOYRALLYDYTAONSDELIDISAGDILAVI-	388	
Db	356	FPAV-----VNGVVEGSGSAGRDLRPRGMFKYQAQHOVDATDIDTDLQAKADVIVLIP	409	
QY	389	-----LEGEDCH-----WTVBERNGO--RGVYRGSTLEKL	415	
Db	410	FONDEBODEGMLVKESSDMNOHKKLECKGSGPPRENTTER	450	

RESULT	11
ID	W94504 standard; Protein; 451 AA.
AC	W94504;
DT	22-APR-1999 (first entry)
DE	Human Blnl protein.
KW	Blnl; brain-specific; box-dependent myc-interacting protein; cancer;
CC	diagnosis; hyperplastic disease; tumour suppressor; gene therapy;
KW	benign prostatic hypertrophy; neurodegeneration.
OS	Homo. sapiens.
PN	W09855151-AI.
PD	10-DEC-1998.
PF	04-JUN-1998; U11647.
PR	06-JUN-1997; US-870126.
PA	(WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI	Prendergast GC, Sakamuro D;
DR	WPI; 99-059881/05.
DR	N-PSDB; X16322.
PT	New splice variant exons of Box-dependent myc-interacting
PT	polypeptide - associated with loss of tumour suppressor activity.
PT	used for diagnosis of cancer and in gene therapy
PS	Example 1; Page 81-82; 13pp; English.
CC	The present invention describes Blnl (Box-dependent myc-interacting
CC	protein 1) brain-specific alternative splice variants exon 12A, B, C and
CC	D. Blnl specific antibodies (when labeled) are used to detect cancers
CC	or other hyperplastic conditions (e.g. benign prostatic hypertrophy)
CC	associated with a deficit of normal Blnl and/or aberrant forms of Blnl.
CC	The same diseases can also be diagnosed at the nucleic acid level using

CC fragments of Bln1 nucleotide sequences in standard amplification and/or
CC hybridisation assays. Cancers that can be detected are carcinomas and
CC epithelial cell tumours, specifically of prostate, liver and colon/
CC rectum, also melanoma. Bln1 nucleotide sequences, when included in
CC a vector, and Bln1 proteins may also be used to treat these diseases,
CC and also degenerative conditions such as neurodegeneration. Bln1
CC proteins may also be used to raise antibodies, and the nucleotide
CC sequences can also be used to express the corresponding proteins. Also
CC anti-idiotypic antibodies can be used therapeutically, and more
CC generally similar materials can be used to treat or diagnose any
CC condition involving deregulation, defect or amplification of the c-myc
CC oncogene. Bln1 nucleotide sequence, proteins, and antibodies are also
CC useful to screen for agents that may be used to treat Bln1-related
CC cancers. The present sequence represents human Bln1 protein.
50 Sequence 451 AA;

Query Match	6.98	Score 149.5	DB 1	Length 451
Best Local Similarity	21.7%	Pred. No. 2.8e-05		
Matches 100: Conservative	70	Mismatches 186	Indels 105	Gaps 20

OY	33	KRMCDVDELLROQAQAERKGLYNQARAGQO--TE----	ANSLTSDLSLKQOTENV	87
Db	17	KLLTAOEKVLQKIGKADETKDEOECVOQNFKNKOLTEGRLOKDLTALASVAKMHE--	74	
OY	88	GSANIQTALALREELRSLEEFEROKOEQRKKEA1-----	MDRQKSKLSYUKTME	139
Db	75	-----ASKINTEGLOEYVEPDMVGRDPAKNIAMNDLMMDYHOK----	LYDQALL	121
OY	140	SKKAT-----DQCRDADDAQAQAEFRSANGHOKOVE----	KSONKAKKCKESATEAERY	150
Db	122	TMDYLLGQFPKIKRIARAGRKLVUDYSARHHYESLQTAKKKEAKTAKAEELIKOKV	181	
OY	191	YKO--NIEOLERARTEWEQENHTTCEAFOL-----	OEFDRLLTILRNALMYHANO----	237
Db	182	FEEEMVNDLOEELPRLMNSRNGVFNNTQSLAGLEENHHEKSKINQMLNDVYLGLEKONG	241	
OY	238	-----LSMOCKYKDELYEAVLTLGECDSVEDINGFTQSKSTGCEHPARVPYQONTYDREV	292	
Db	242	SNTEFTVKAQOPRKSKFLRLRRKKNSDNAPAKGN--	KSPSPDGSAPATR--ETRVNHER	297
OY	293	TPLLIS-----SPS-----	IOPSCVILKRFSGLLHGSKTPRSPAASTET	332
Db	298	EPAGATGATGALPSPSPQAPASEVAGSTOPAGA--	QEPQETPLASAPASSSLPAAVYET	355
OY	333	LTPPERNELLYVASIEVQATOGNUN--	SSAODYRALUDYTAONSDELIDISAGDILAVI-	388
Db	356	FPAT-----VNGIVEGSSGAGRLDRLPRGFEMFYQAQOHADYATDPTDLOKADGVYLYIP	409	
OY	369	-----LEGEDGW-----	WTVERNGO--KGFVNGSTLEKX	415
Db	410	FQNPDEQEGWIMGVKESDMNQHKKLKCRCORPPENTRETEV	450	

RESULT	12
ID	W89721
AC	W89721 standard; Protein; 1484 AA.
DT	W89721;
DE	21-MAY-1999 (first entry)
KW	Canine ribosome receptor.
OS	Ribosome receptor; dog; protein secretion; eukaryotic cell.
PH	Canis familiaris.
FT	Location/Qualifiers
FT	Key
FT	Peptide
FT	1..28
FT	/note="membrane anchor"
FT	29..736
FT	/note="basic N-terminal region"
FT	737..1535
FT	/note="acidic C-terminal region"
FT	157..736
FT	/note="54 repeats of 10-amino acid consensus motif"
PN	MO9901565-A1.

PD 14-JAN-1999.
 PF 02-JUL-1998; U13870.
 PR 03-JUL-1997; US-051721.
 PA (RECC) UNIV CALIFORNIA.
 PI Becker F, Meyer DI;
 DR WPI: 99-106065/09.
 PT Increasing secretion of proteins in eukaryotic host cells - using a
 PT host genetically altered to express increased ribosome receptor
 PS Disclosure; Fig 1A; 39pp; English.
 CC This is the amino acid sequence of the 180 kDa canine ribosome
 CC receptor, as deduced from a cDNA sequence. The ribosome receptor
 CC induces membrane proliferation in various eukaryotic cells.
 CC resulting in increased secretory capacity in these cells. A
 CC claimed process for increasing the intracellular transport or
 CC secretion of a desired protein (PI) in a eukaryotic host cell
 CC comprises: (i) obtaining a eukaryotic host cell, genetically
 CC altered to express increased ribosome receptor or its fragment; and
 CC (ii) expressing PI in that cell so that transport or secretion of
 CC PI is increased with respect to that in the wild-type host. Also
 CC claimed is a method for enhancing production of a recombinant
 CC protein comprising: (a) isolating a recombinant DNA molecule
 CC encoding at least a portion of a ribosome receptor; and (b)
 CC introducing the recombinant molecule into a eukaryotic host cell to
 CC form a genetically engineered cell so that the cell is capable of
 CC secreting increased amounts of a recombinant protein when its
 CC coding sequence is expressed. The process can be used in the
 CC production of secreted proteins, such as therapeutic proteins, e.g.
 CC tissue plasminogen activator or erythropoietin. The invention
 CC provides a more economical means to produce proteins in eukaryotic
 CC cells than prior art.
 SQ Sequence 1484 AA.

Query Match 6.9%; Score 149; DB 1; Length 1484;
 Best Local Similarity 20.9%; Pred. No. 0.0002;
 Matches 95; Conservative 68; Mismatches 177; Indels 114; Gaps 16;

QY 33 RKCKDVEELL-RORAOEERYKELVOIARKAGGOTEMSLRTSPDSLKOETENVSAN 91
 DB 899 REHVKEVOOQLOGKIRTLQEOLENGPNTOLARL--QOENSLTR--DALNQTQSVESKQ 952
 QY 92 IOLALIREELRSI-----EEREROKROKKYEALIMRVOKSKLSYKKTMEKAY 144
 DB 953 NTELAKLROELSKYKVELKSEAROEEOORALETKTALAEKQVLOLQASHKESEAL 1012
 QY 145 DQR-----CR-----DADDAEQAFERVSANGHOKQVEKSONKAK----- 178
 DB 1013 QKRLDEVSRRLCSQTSASHLRADAKEAQOQOMALHSLKLOSSEAEVSKSELSGLH 1072
 QY 179 -QCKESATEERYRONIDQLERARTEWEOHRTTCEAFOLQEDRLTILNALWVHC-- 235
 DB 1073 GQLEKARAEKNSOL-MERIRSIETALLLEAGQARDTQDAQASRAEHQARKLEESQVW--CLE 1129
 QY 236 -----NOLSMQCVKDELLEYEVRLTLE-----GCDVEGDIINGTQSKSTREPPAP 281
 DB 1130 KEATTELKAEVQOKVKNNDREKNKAMEALASERAC--EKKRLSTLQKKESEK----- 1183
 QY 282 VPYONYVYDREVTPLIGSPISQPCGVIKRPSGLHSGPKTTPSPAPASTETLTPPTPER-- 339
 DB 1184 -----QLSITEAQOTKEALLLALLPALSSASAPSYTEMIGELREKGP 1223
 QY 340 -----NELVYASIEVQATOGNINSSAQDYRALYDYTAONSDELISAGDILA 386
 DB 1224 ELTKORPADTDPSSDLASKRLAEAEFTQNNIQAECDDQYRTILAEETGMLKDLQKS----- 1277
 QY 387 VILEGEDGMW-----TYERNQGGVPGSYLEKL 415
 DB 1278 --VEEEDQVWKAQVSAEELQKSRVTKHLEDI 1309

RESULT 13
 W05399
 ID W05399 standard; Protein; 509 AA.

AC W05399;
 DT 19-FEB-1998 (first entry)
 DE Human clone 65 protein.
 KW Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
 KW cellular signalling element; cellular structural element; malignancy;
 KW protein identification; functional domain; protein screening;
 KW cellular signal transduction process.
 OS Homo sapiens.
 PN W09631625-A1.
 PD 10-OCT-1996.
 PF 04-APR-1996; U04454.
 PR 03-APR-1996; US-630915.
 PR 07-APR-1995; US-417872.
 PA (CYTO-) CYTOGEN CORP.
 PA (UYNC-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;
 DR WPI: 96-465045/46.
 DR N-PDB: T39799.
 PT Identifying polypeptide(s) having specific functional domain (esp.
 PT SH3 domain) - comprises detecting selective binding to recognition
 PS unit, regardless of sequence homology
 PS Claim 54; Fig 59; 174pp; English.
 CC W05386-W05403 represent novel human and mouse Src-homology region 3 (SH3)
 CC domain containing proteins that can be used in the method of the
 CC invention. SH3 domain containing proteins play a role in signalling and
 CC structural elements of cells. The method of the invention is for
 CC identifying polypeptides containing functional domains of interest
 CC (especially SH3 domains). The method comprises contacting a multivalent
 CC recognition unit (RU) complex with a number of peptides and identifying
 CC polypeptides having a selective binding affinity for the RU complex. The
 CC method is based on functional similarities and does not rely on sequence
 CC similarities. Prior methods only gave limited success for identifying
 CC proteins which contain an SH3 domain due to the minimal sequence homology
 CC among known SH3 proteins. It has been found that small peptide RUs in
 CC multivalent form have reduced specificity for a given functional domain
 CC compared to monomer RUs. Multivalent RU complexes are particularly suited
 CC to screening for polypeptides containing functional domains that are
 CC similar to, but not identical in sequence to, the original target
 CC functional domain. The new method enables proteins having a common
 CC function to be identified. Identification of novel SH3 proteins will be
 CC useful for a better understanding of cell growth, malignancy, signal
 CC transduction processes, etc. New candidate drugs can be identified, and
 CC their specificities (e.g. pharmacological activities) can be assessed
 CC using the method of the invention.
 SQ Sequence 509 AA;

Query Match 6.9%; Score 148.5; DB 1; Length 509;
 Best Local Similarity 24.1%; Pred. No. 4.2e-05;
 Matches 76; Conservative 45; Mismatches 116; Indels 79; Gaps 14;

QY 126 VOKSKLSLYKKTMEKKRAYDQKCRDADDAEQAFERVSANGHOKVEKSONKAKCK----- 181
 DB 5 IERRKLEL-----WQKKLLEDEAR--KAKQGENLMKELRKREBEKQKRLQOEKTOEK 57
 QY 182 -ESATEAEVRYRONIDQLERARTEWEOHRTTCE--AFOLQEDRLTI----- 226
 DB 58 IOEERKAEKQKETAASVLYNRYALYFEARNHDEMSFGDIIQVDEKTPGPMWLYGS 117
 QY 227 -LRNALWVHCNOLSMQCVKDELLEYEVRLTLECGDVEGDIINGTQSKSTREP-----PA 260
 DB 118 FQGNFGFPCNVYKEMSSSEKAVSPKALLPPTVS-----LSATSTSSSEPLSSNOQA 171
 QY 281 PV-PYONYVYDREVTPLIGSPISQPCGVIKRPSGLHSGPKTTPSPAPASTETLTPPTPER 339
 DB 172 SVTDYQNV-----SESNL--TYNTSWQKSAFTRTVSP----- 202
 QY 340 NELVYASIEVQATOGNL--NSSAQDYRALYDYTAONSDELISAGDILAVILEGEDGMWT 397
 DB 203 -----GSVSPITHGGQVVENLKAQ---ALCSWPAKKADNHLNFSKHILITY-LDQGNWME 253
 QY 398 VERNQGRGVPGSYLE 413

Db 254 GEVHGGRGMPKSYK 269

RESULT 14

W82799 standard; Protein: 534 AA.

AC W82799; 10-MAY-1999 (first entry)

DE Human IRS-p53h protein.

KW IRS-p53h; Insulin receptor tyrosine kinase substrate; treatment;

KW Insulin response; inflammation; cancer.

OS Homo sapiens.

PN W09857987-A2.

PF 23-DEC-1998.

PE 19-JUN-1998; U12903.

PR 19-JUN-1997; US-878563.

PA (INCY-) INCYTE PHARM INC.

PI Hillman JL, Lal P, Shah P;

DR WPI: 99-080949/07.

DR N-PSDB: V72220.

PT New insulin receptor tyrosine kinase substrate - useful for treating

PS Insulin-related disorders, inflammation and cancer

PS Claim 1; Fig 1A-F: 66pp; English.

CC This sequence represents a novel human insulin receptor tyrosine kinase substrate (IRS-p53h). IRS-p53h or its agonist is used to treat a disorder associated with insulin response, and its antagonist is used to treat inflammation and cancer. Particular insulin response disorders include type 2 diabetes mellitus, hyperglycemia, myotonic muscular dystrophy, acanthosis nigricans, retinopathy, nephropathy, atherosclerotic coronary and peripheral arterial disease, and peripheral

CC and autonomic neuropathies. Inflammation disorders include adult respiratory distress syndrome, allergies, asthma, bronchitis, emphysema, hyperosinophilic, myocardial or pericardial inflammation, rheumatoid arthritis, Addison's disease, AIDS, anaemia, atherosclerosis, various

CC diseases of the digestive system, atopic dermatitis, dermatomyositis, diabetes mellitus, glomerulonephritis, gout, Grave's disease, lupus erythematosus, multiple sclerosis, myasthenia gravis, osteoarthritis, osteoporosis, pancreatitis, polycystic kidney disease, polymyositis, scleroderma, Sjogren's syndrome, autoimmune thyroiditis, complications

CC of cancer, extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal and helminthic infections, and trauma. Antibodies specific for IRS-p53h may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical

CC agents to cells or tissue which express IRS-p53h.

CC Sequence 534 AA;

SQ

Query Match 6.9%; Score 148.5; DB 1; Length 534;

Best Local Similarity 21.6%; Pred. No. 4.5e-05;

Matches 95; Conservative 72; Mismatches 177; Indels 95; Gaps 21;

QY 35 MCDVDELLRQRAQAEERKELY---QIARRKGGOTEMSLTSPSLKQOTEN----- 86

Db 34 MGNRYKALAGTYAAKGFADLVKMGELASEQSKELDVLFQNAEVRQIONOLEEM 93

QY 87 VGSANITQALALREELRSEEF-----RERQKQKRYEAIMDRVQSKSLYKKTWESK 141

Db 94 LKSFHNEILLTQLEQKVELDSRYLSAALKKRYQTRQSGDA-LDKQALKKLKRKSGSK 152

QY 142 ---KAVDQKCRAD---DDEQAFERYVSAHQKQVKSQK-----AKCKESATEAERV 190

Db 153 NPQKYDKELOYIDAI SNKGELENVSDGKPTALTEECRRFCFLVEKQCAVAKNSA-AV 211

QY 191 YRONISQLEBARTMEQEHRTTCEAFQLOEFDELITLRNALWGHONLSMQCVKDDLYE 250

Db 212 HSKGKELLPOKPLMOQ---ACA-----DPSKIPERAV-----OLMOQVANSNG--- 251

QY 251 EVRLTLEGCDVEDINGFIQSKSTGEPPAPVYQNTYDREVTPLGSPISQSCGYIKR 310

Db 252 ---ATLPALASAKSNLYISDPIPAK-PLPVP-----PELAPFVGMSAOSTPIIMG 301

QY 311 FSGLLHGSPKTPSAP-----AASTELTPTPRNET-----VYAST 347

Db 302 VTG-----PDGEDSPWADRKAQPKSLSPQSQSKLSDSYNTLPVRKSYTPRNSYATT 356

QY 348 EVQATQGNLNSA-----QDYRALYDYTA-QNSDELISAGDILAVIL-EGREGW--W 396

Db 357 ENKTLPRSSMAAGLERGRMRVAKAFISHAAGDNSTLSFEKGLITLLVPEARDGMYG 416

QY 397 TVERNGQGFVPGSYLEKL 415

Db 417 ESEKTKMRGMPFSYTRVL 435

RESULT 15

R47173

ID R47173 standard; Protein: 2101 AA.

AC R47173;

DE 15-JUN-1994 (first entry)

DE Sequence of the inner nuclear matrix (INM) protein MT2.

KW Malignant cell type marker; interior nuclear matrix; MT2; NuMA.

OS Homo sapiens.

PN W09400573-A.

PD 06-JAN-1994.

PE 21-JUN-1993; U06160.

PR 22-JUN-1992; US-901701.

PA (MATR-) MATRITTECH INC.

PI Lidgard GP, Toukally G;

DR WPI: 94-026210/03.

DR N-PSDB: 054841.

PT Novel malignant cell type markers of the interior nuclear matrix

PT - used for detecting abnormal cell types e.g. malignant breast,

PT prostate, lung, etc., for determining deg. of cell death in

PT tissue, etc.

PS Claim 39; Page 56-66; 93pp; English.

CC MT1 and MT2 proteins were isolated from malignant cells essentially following the method of Penman and Fey, described in US Pat. Nos. 4, 882,268 and 4,885,236. DNA sequences encoding these proteins were

CC raised against the isolated INM proteins. The genetic sequence encoding MT2 has been disclosed by Yang et al. 1992, J. Cell Biol. 116:1395-1408, and is referred to by them as 'NuMA'. The nucleotide

CC sequence encoding MT1 has not been described previously. Both the MT1 and MT2 proteins are distributed throughout the nucleus (with the exception of the nucleolus) in non-mitotic cells, and localise to the spindle during mitosis. A binding protein having a binding affinity of greater than about 10(9) M(-1) for R47173 is claimed

CC for use in the manufacture of a medicament.

SQ Sequence 2101 AA;

Query Match 6.9%; Score 147.5; DB 1; Length 2101;

Best Local Similarity 22.1%; Pred. No. 0.00046;

Matches 101; Conservative 63; Mismatches 146; Indels 147; Gaps 22;

QY 19 HTGEVILLRLDGRKMKCVDELLRQRAQAEERKELYQIARRKAGG-----QTEM 70

Db 1350 HTSQALVSEILPAKHLC---QQLQAEQDAEAKHRELEQSKQAAGLRAELRAQREL 1406

QY 71 -----NSLRTSPDSLKQOTFENNNGSHIQIALALREBLRLEEF----- 107

Db 1407 GELIPLRQKVAEBERTAQOQLAEKASTAEQLSMKKHGLA-----EBNRQGERANIGR 1462

QY 108 -FRERQKQ-RKKY-----EAIMDRVQSKSLYK-----KTWES 140

Db 1463 QPLEVELDQAREKYVQELAAVRAADAEIRLAQVQREADSTARELVMTRAKYGCAYKVYLE 1522

QY 141 KKAAYDQ-----KCRDADA---EQAPEKVSANG---HOKQVKSQNKAKQC 180

Db 1523 RQRPQEROKLITQAQVELSKKLADSDASKVQOQKLAQVQAGGSEQDEARFOAQLNEL 1582

QY 181 KESATFAEARYVRQNIQELERARTEWE-----QEHRTTCEAFQLO-----EPRRLT 225

Db 1583 QAOISQKEQAAEHYKIQIMEKAKTHYDAKKQONQELQRLQLEQKELRAEABERLG 1642

QY 226 ILRNALWGHQNSMQCVKDELYEVR-LTLECGDVEGDI-NGFIQSKSTG----- 275

```

D6 1643 -----HELDQAGLKTKEAEQFCRHHLTQAVRSELEQVAHHADQDLRLDGLKFCQVATDA 1692
QY 276 ---REPAPVPYQNYD-----REYPLTGS--PSIDSCGVAKRFGSLHGSPKTTT 323
D6 1693 LKSRPEPAKPOLDLSIDSLDLSCEEGTTPSLTISKUPLRTP-----DGTSTVP 1738
QY 324 SAPAAS--TETLTPPERNELVYASIVQATQGLNLS 359
D6 1739 GEPASPIISQRLPRVSELESLYTTPLPARSQAPLES 1775

```

Search completed: September 15, 2000, 10:36:50
Job time: 1984 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2000, 10:24:18 ; Search time 48.22 Seconds
(without alignments)
532.604 Million cell updates/sec

Title: US-09-068-377-1

Perfect score: 2152

Sequence: 1 MMAOLQFRDAFWCRDFAHT.....WTVERNRGQRFVPSYLEKL 415

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_64:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	346	16.1	670	2 S67383	probable signal tr
2	345.5	16.1	927	2 T38127	phosphoprotein - f
3	325	15.1	858	2 S54119	phosphoprotein - f
4	325	15.1	907	2 A57087	cell division cont
5	283	13.2	906	2 T00039	hypothetical prote
6	199.5	9.3	1011	2 T13055	dynamlin associated
7	199.5	9.3	1094	2 T13053	dynamlin associated
8	194.5	9.0	1270	2 T09194	adaptor protein in
9	191.5	8.9	416	2 B48580	antigen EG13 - tap
10	185.5	8.6	2427	2 T16613	hypothetical prote
11	177.5	8.2	426	2 A48880	echinococcal micro
12	177	8.2	501	2 T39801	hypothetical sh3-c
13	176.5	8.2	1113	1 A47106	myosin heavy chain
14	173	8.0	783	2 T23452	hypothetical prote
15	171	7.9	1168	1 T23456	myosin heavy chain
16	170	7.9	785	2 T23456	hypothetical prote
17	169	7.9	1099	1 S31926	myosin IB heavy ch
18	167	7.8	546	2 I48899	coractin - mouse
19	166.5	7.7	633	2 S48956	hypothetical prote
20	166.5	7.7	1736	2 T00391	hypothetical prote
21	166	7.7	550	2 A48063	mammary tumor/squa
22	165	7.7	526	2 A34855	67k neutrophil oxi
23	165	7.7	816	2 T17257	hypothetical prote
24	163.5	7.6	2477	1 SJC4A	spectrin alpha cha
25	161.5	7.5	496	2 JCS521	salt-tolerant prot
26	161	7.5	2472	2 A35715	fodrin alpha chain
27	160.5	7.5	642	2 T39376	hypothetical DAG-b
28	159.5	7.4	482	1 S40887	RVS167 protein - y
29	159.5	7.4	1097	2 T31504	hypothetical prote

30	159.5	7.4	1109	2 S53601	myosin-IC - human
31	158.5	7.4	1459	2 T30196	kinesin motor prot
32	158	7.3	563	2 A41530	src substrate p80/
33	158	7.3	1920	2 A53188	pericentrin - mous
34	157.5	7.3	1107	1 S52517	myosin I heavy cha
35	153.5	7.1	482	2 JCS593	amphiphysin II2 -
36	151.5	7.0	454	2 S01092	fodrin alpha chain
37	151.5	7.0	1133	2 T22976	hypothetical prote
38	151.5	7.0	2415	1 A33733	spectrin alpha cha
39	150.5	7.0	769	2 T22316	hypothetical prote
40	149	6.9	475	2 T16364	hypothetical prote
41	149	6.9	1534	2 A56734	ribosome receptor,
42	147.5	6.9	2101	2 A42184	nuclear mitotic ap
43	147	6.8	1132	2 T00259	hypothetical prote
44	146.5	6.8	466	2 T11684	RVS167 protein hom
45	146	6.8	2442	2 T08621	centrosome associa

ALIGNMENTS

RESULT 1

Probable signal transduction protein, FCH domain and SH3 domain containing - fission
C/Species: Schizosaccharomyces pombe
C/Date: 19-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 31-Jan-2000
C/Accession: S67383; T39317
R:Murphy, L., Harris, D.
Submitted to the EMBL Data Library, February 1995
A/Reference number: S67376
A/Accession: S67383
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-670 <MUR>
A/Cross-references: EMBL:Z69379; NID:q118396; PID:e221942; PID:q1184004
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
Submitted to the EMBL Data Library, September 1998
A/Reference number: Z21844
A/Accession: T39317
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-670 <WOO>
A/Cross-references: EMBL:AL031528; PIDN:CAA20684.1; GSPDB:GN00067; SPDB:SPBC11C11.02
A/Experimental source: strain 972h-, cosmid c11c11
C/Genetics:
A/Map position: 2
A/Map position: 2
A/Introns: 23/3; 54/2; 612/1
C/Superfamily: SH3 homology
F:614-665/Domain: SH3 homology <SH3>

Query Match 16.1%, Score 346; DB 2; Length 670;

Best Local Similarity 18.6%, Pred. No. 1.4e-13;

Matches 123; Conservative 84; Mismatches 191; Indels 264; Gaps 13;

QY	7	FRDAWCDFTAHNGEYVLORLIDGRKCKRVEELRORAAEERKGEIVQIRKAGG	66
DB	18	FSNFWGAM---DEGYHLLSFSFDVKHINEELRBFYHERANIEDYAKRKAKLSRTTFS	74
QY	67	QTEMSLTFSDSLKQOTENVSANHQLALREEL-RLSEFFRRORQRKKYEAIDMR	125
DB	75	SLEGLCKESYQVMKAEDVNMKSHLQSLLODDVENAFTRYASLKDKKMIYSGIEK	134
QY	126	VQKSLSLYIKKTMEKRAKYDQKCRADAEOAFERYVANG--HQQVEKSKAKACKES	183
DB	135	VHKRLSKHQALVNAQDDYHYLCKRVN-----YVSOQNMLEGELEKNNKLNKTQNA	188
QY	184	ATEAEVRYRONIEOLERARTEWDEHRTCEAFQLOEFDRLTILNNALVHNCQSMOCV	243
DB	189	ITASSSDQSAVNAVRDYSARKNTNEMRSTCDKLDQIEERKRLFLKSVMTTLLISRGCF	248
QY	244	KDDELYEEVRLTLRGCDVEGDINGFIQSKSTGREPPAPVPYONYVDREV-----	292

```

Db      249 NNDQACERIRKMLEQCSVSQVLEFLIDAKSTGTGIPPKFYDYKGEVDDSVELYQAN 308
QY      293 -----
Db      309 FQRAQTKIENDMPLNRPYVLASATARNESPENTLPNTSAIQSLTIVSSNSQNGSSP 368
QY      293 -----TPDIG-----SPSIQSCGVYIKRRSGLLHSGPK--TT 322
Db      369 KKSFLSKPKLTSRSTPVGWGTAPDALSSPRNDSPPLTSADEQMKHLIS--LOBEPKONPT 426
QY      323 PSAPAASTETLTPTPERNEL-----IEVATQGNL----- 356
Db      427 PAAGAFPNSTLTLPRLNELGSLFPNSVSFTEDSRPNVTPSRROQIOEFEGSVLQMEN 486
QY      343 -----VVAS----- 356
Db      487 RAVSPVDSKNGSRSSFTLRKSRSPRPSSLSQNASRLPRSLTIPGNLEPNYDFGVVD 546
QY      357 -----NSSAOD-- 362
Db      547 PASGTAPTDEPTDSSFVDDTINTKATGNTSNRLSLPAYPTDGGDSTIDNPTSTGQ 606
QY      363 -----TRALYDTAQNSSDELIDISAGDILAVILEGEDGMW---TVERNGQGFVPGSTLE 413
Db      607 RILGYSAALYDAAIPEIISFRKGGDTIATVLTLYEDGMWEGFVGEDDHNGGFPNSFVR 666
QY      414 KL 415
Db      667 EI 668

```

RESULT 2

phosphoprotein - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T38127
 R:Badcock, K.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, May 1997
 A:Reference number: 221772
 A:Accession: T38127
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-927 <BAD>
 A:Cross-references: EMBL:295334, PIDN:CAB08599.2, GSPDB:GNO0066, SPDB:SPAC2068.05c
 A:Experimental source: strain 972h-, cosmid c2068
 A:Genetics:
 A:Gene: SPDB:SPAC2068.05c
 A:Map position: 1
 A:Introns: 27/3; 58/2; 871/1

Query Match 16.1%; Score 345.5; DB 2; Length 927;
 Best Local Similarity 25.98; Pred. No. 2.2e-13;
 Matches 96; Conservative 74; Mismatches 169; Indels 31; Gaps 8;

```

QY      3 AQLQRDFAWCDPFAHTGYEVLQRLDGRKCKDVEELLQRAQAEEYRGKELVQIAR 62
Db      18 SSVKFRDNWQSE--DAGMDALMSRTSSLSVLESIDEFYAKRASIEREYASLQELAA 74
QY      63 KAGGQTEVNSLRTSPDLSKOOTENVGSAHIQALALREELRS--LEEFREKQKQRYEA 121
Db      75 SSAIPEVGSLLNLLSKRTETGSMARAHVEVSQOINTELRNKIRREYIDQTEQKVVAA 134
QY      122 IMDRVQSKSLVYKRTMESKAYDQCRDADAQAEFRYSANGHOKOVEKSONKAKQCK 181
Db      135 AIEFLYKQKTALEIDISEKKDAVEYSCNKLNSTYMKTKMTG---RELDKYNLKITRQAA 190
QY      182 ESAFEARVYRONIEQLEKARTQEWQDHRTTCEAFQLOEFDRLLTIRNALWVHCNQLSMQ 241
Db      191 LAVKKDAEIRETNELLLTVTRWEMIDRWTEVCDAFHIEYRLEFLKTMAYANITSTA 250

```

```

QY      242 CVKDELYEVRLLTEGCDVEGDINGFIQSKSTGREPAPVYONYDRE-----VTPL 295
Db      251 CVKDESCERKIRLTLENINIDITQIONEGTGTTP-PLPEFNDYFKENGLNYDIDLQ 309
QY      296 IGSPSIQSCGVYIKRRSGLLHSGPKTTPSAPAASTETLT-----PT-----PERN 340
Db      310 ISKASVYFSSSRPSASLASLSPRSFRAFRPKTS-ETVSSVEVSSPTTSLHSPKAVSN 368
QY      341 ELVVASIEVQ 350
Db      369 EOVEQEVEVE 378

```

RESULT 3

S54119

phosphoprotein - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 25-Oct-1996
 C:Accession: S54119
 R:Fankhauser, C.; Raymond, A.; Cerutti, L.; Simanis, V.
 submitted to the EMBL Data Library, April 1995
 A:Description: The cdc15 gene is a key element in F-actin reorganisation at mitosis.
 A:Reference number: S54119
 A:Accession: S54119
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-858 <FAN>
 A:Cross-references: EMBL:X86179
 A:Genetics:
 A:Introns: 38/2
 C:Keywords: phosphoprotein

Query Match 15.1%; Score 325; DB 2; Length 858;
 Best Local Similarity 25.98; Pred. No. 3.4e-12;
 Matches 91; Conservative 70; Mismatches 163; Indels 28; Gaps 7;

```

QY      21 GYEVLLQRLDGRKCKDVEELLQRAQAEEYRGKELVQIARFAGQTEVNSLRTSFDL 80
Db      13 GMDALMSRTSSLSVLESIDEFYAKRASIEREYASLQELAASSADIPEVGSLLNNTISM 72
QY      81 KOQTEVNSAHIQALALREELRS--LEEFREKQKQRYEAIMDRVQSKSLYKKTWE 139
Db      73 RTEGSAKAAHEEVSQOINTELKRNKIREYIDQEQKVVAAANIEFLYKQKTALEIDISE 132
QY      140 SKRAYDQCKDADADAQAEFRYSANGHOKOVEKSONKAKQCKESAFAEARVYRONIEQLE 199
Db      133 KDAIYETCKNLNSYMRQTKMTG---RELDKYNLKITRQAAIAVKKDAEIRETNELLL 188
QY      200 RARTEWQEHRTTCEAFQLOEFDRLLTIRNALWVHCNQLSMQCVKDELYEVRLLTEGC 259
Db      189 TVTRWIDRWTEVCDAFHIEYRLEFLKTMAYANITSTACVADDESCRIRLTLENT 248
QY      260 DVGVDINGFIQSKSTGREPAPVYONYDRE-----VTPLIGSPSIQSCGVYIKRRFSG 313
Db      249 NIDEDITQIONEGTGTTP-PLPEFNDYFKENGLNYDIDLISKASVYFSSSRPSASA 307
QY      314 LHSKPTTPSAPAASTETLT-----PT-----PERNELVVASIEVQ 350
Db      308 SLASSPTRSAPFRKTS-ETVSSVEVSSPTTSLHSPKAVSNQVEQEVEVE 358

```

RESULT 4

A57087

cell division control protein cdc15 - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 07-May-1999
 C:Accession: A57087
 R:Fankhauser, C.; Raymond, A.; Cerutti, L.; Utzig, S.; Hofmann, K.; Simanis, V.
 Cell 82, 435-444, 1995
 A:Title: The Schizosaccharomyces pombe cdc15 gene is a key element in the reorganizat
 A:Reference number: A57087; MUID:95360987
 A:Accession: A57087

A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-907 <FAN>
A:Cross-references: GB:X86179
C:Genetics:
A:Gene: cdc15
A:Insertions: 38/2, 851/1
C:Superfamily: SH3 homology
C:Keywords: mitosis
F:853-904/Domain: SH3 homology <SH3>

Query Match 15.1%; Score 325; DB 2; Length 907;
Best Local Similarity 25.9%; Pred. No. 3,7e-12;
Matches 91; Conservative 70; Mismatches 163; Indels 28; Gaps 7;

21 GYEVLLQRLLDGRKMCDEVELLRQRAQAEERKELVQIRKAGGQTEMSTLSPSL 80
13 GMDALMSRTKSSLSVLSIDSEDFYAKRASIERYASKLOELASSADIPVEGSTLNILSM 72
81 KOOTENGSAHIOLALALREELRS-LEEFREKOEKRYKPAIMDRVQSKSLSYKKTME 139
73 RTETGSAKAKHEEVSQOINTELKRIKREYIDQEOKVANALIELYOKTALFIDISE 132
140 SKRAYDQCRDADDAEQAFERVSANGHOKOVESONKAKOCKESATEAERYRONIEOLE 199
133 KRDAYESCCKNLSYMKQTKMTG---RELDKYNLKIQAALAVKKMDAERYETNELLL 188
200 RARTEMQEHRTTCEARLOLEFDRLTLIRNALWCHNLSQCKYKDELEYEVNLTLEGC 259
189 TWTREWDRTMEYCDVQHEERLEFLKTNMAYANITSPACVKDESCERIKRLTLENT 248
260 DVEGDINGFIQSKSTGREPPAPVYQNYDRE-----VPLIGSPSTIQPCGYIKRPSG 313
249 NIDEDITQMNQNETGTTIP-PLDEFNDYKRENLNDIOLIKABSPYSSSRPSASA 307
314 LHHGSKPTPSAPASTETLT-----PT-----PERNELVYSIEVO 350
308 SLASSPFRSAFRRPKTS-ETVSSSEYVSSPPTSPLHSPVKPVSNEQVEQTEVE 358

RESULT 5
T00039
hypothetical protein KIAA0290 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 05-Nov-1999
C:Accession: T00039
R:Ohara, O.; Nagase, T.; Ishikawa, K.; Nakajima, D.; Chira, M.; Seki, N.; Nomura, N.
submitted to the EMBL Data Library, August 1997
A:Description: Prediction of the coding sequences of unidentified human genes.
A:Reference number: Z14073
A:Accession: T00039
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-906 <OHA>
A:Cross-references: EMBL:AB006628; NID:d1170681; PIDN:BAA22959.1; PID:d1023835
A:Experimental source: brain
A:Genetics:
A:Note: KIAA0290

Query Match 13.2%; Score 283; DB 2; Length 906;
Best Local Similarity 24.3%; Pred. No. 1,2e-09;
Matches 93; Conservative 90; Mismatches 148; Indels 52; Gaps 15;

7 FRAFAFCRDTAHTGYEVLLQRLLDGRKMCDEVELLRQRAQAEERKELVQIRKAGG 66
21 FGCHFWGE---KHHGEVLHYSKQGIPTKELADIFREARTIEFYSKAKAKSLKLSN 77
QY 67 QTEMNLSRTSFDLSKOOTENGSAHIOLALALREELRSLEEFREKOEKRYKPAIMDRVQSKSLSYKKTME 122
DB 78 GTMGKTFAPRLMEYFRVSSDKLALCHLELTRKLODLINKVLRGEEQLKTHKKCKEEVAVST 137

QY 123 MDRVQSKSLSYKKTME-ESKAYVDQCRDADDAEQAFERVSANGHOKOVESONKAKOCK 181
DB 138 IDAVO--VLSGSQQLPKSRKENVLNCMD---QELRLREESTS--QKEMDKAKETTKKKA 189
QY 182 ESATFAERYRONIEOLEARTMEQEHRTTCEARLOLEFDRLTLIRNALWCHNLSQCKYK 241
DB 190 ES-----LRSVERKYNARADFEQKMDSALRFQAMEETHLRHMKALLGSVAHSEVT 242
QY 242 CYKDELEYEVNLTLEGCVEGDINGFIQSKSTGREPPAPVYQNYDREVPPLIGSPST 301
DB 243 HVQIQVHEEFKQNIENSVEMILKRFASKGTGKRPGLPEAY-----SAAL 293
QY 302 QPSCGVKRFSGL---LHG-SPKTPPSAPASTETLTTP---TPERNELVYSIEVOAT 352
DB 294 QCA---MKRLRCAKAPRLPGLSREREPRPAPVPLEBDSGCEPVEDEGF-TVAPDVT 349
QY 353 QGNLMSADYALDYTAQNSD 375
DB 350 Q---NSTAEPSP---FSSSDSD 365

RESULT 6
T13055
dynamlin associated protein isoform 160-2 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999
C:Accession: T13055
R:Roos, J.; Kelly, R.B.
J. Biol. Chem. 273, 19108-19119, 1998
A:Title: Dap160, a neutral-specific Eps15 homology and multiple SH3 domain-containing
A:Reference number: Z17594
A:Accession: T13055
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1011 <ROO>
A:Cross-references: EMBL:AF054612; NID:g2996029; PID:g2996030; PIDN:AAC39139.1
A:Genetics:
A:Gene: Dap160

Query Match 9.3%; Score 199.5; DB 2; Length 1011;
Best Local Similarity 22.6%; Pred. No. 0.00014;
Matches 95; Conservative 80; Mismatches 173; Indels 73; Gaps 18;

QY 33 RKMCDVEELLRQRAQAEERKELVQIRKAGGQTEMSTLSPDLSKOOTENGSAHI 92
DB 348 RKIMD-----QQRKEEREREKREADKREKARLEAE--RKQDELFRLQROREIEM 400
QY 93 QALALREELRSLEFR--ERQEO-----RKKYEALIMDRVQSKSLSY 134
DB 401 EKEEQKRELEKKAARKELEKQROQEWQARIAEWNQKKEQERVLKQKANHQLANE 460
QY 135 KKTMSK-KAYDQCRDADDAEQAFERVSANGHOKOVESONKAKOCKESATEAERYRON 193
DB 461 LSTLEKIKELSORICDPAGVTNVKTV-IDGMRTQORDTSMEGOLK-----ARIKQ 513
QY 194 NIE--QLRARTMEQEHRTTCEARLOLEFDRLTLIRNALWCHNLSQCKYK--DELY 249
DB 514 NKLQDLTQERKAKESKASGALGENAOEOI--NAAFN-RKLIINQIKVIENTS 570
QY 250 EEVRLTEGCBVEDIN-GFIQSKSTGREPPAPV-----YONYVDREVTPLIGSPST 303
DB 571 KEIESKE-----DINNDVGMSELKAKELSLIKCELYKEY-----DYOR 612
QY 304 SCGVTKRPSGLHSGPKTTPSAPASTETLTTPPERNELVYSIEVOATQ--NINSSA 360
DB 613 TSVLEIKYRNKNETSVSSAMDTGSSAMEETGTVTDPAYASNDISALAPAYDLGPA 672
QY 361 QD---VALDYTAQNSDELDISAGDILAVILE--GEDGMWTVERNQGRGVPSSTYK 414
DB 673 PEGFYQAVYEFNARNAEITFVPGDITILVPLEONAEPSGLAGINQHTGFPSYER 732
QY 415 L 415

Qy	93	QIALALREER	-----	SLEPPE----	RQEDRKKYEAIMRVOKSKILTKKTMESK	142
Db	56	NIAAGLINPVOLGIKMKORENFHSSISTIS	IKKVNKFNDSFEENQAQWTYKHKKNVRCK	115		
Qy	143	AYDOCKRADDAEQAFERVENANGHOKOVEKSONKAKQCKESATEAREVYRONTEOLERAR	202			
Db	116	EYFHACK-----	TVRSLOVYOYAKNKEPPGTPEQAQRCKELKMKEDLRGI	163		
Qy	203	TEWEOHRTTCEAF-----	QLOEPRLTLIRNALVHNCJMSOCV	243		
Db	164	MEEEKTRKAYEAELSSLDVTPRIEDMTGVFNKAQAFERERI----	YVFKQALOMQAV	219		
Qy	244	KD-----	DELVEVALLTEGCDVEBDINGFIQKSTGREGPAAPYONYNDREVTPLI	296		
Db	220	LDISAKPNLSOIFVELRETYAKADADADLKKMSIAGVDMAPFVY-FOEY-----	269			
Qy	297	GSPSIQ-----	PSCGV----IKRFSGLHGS--PPTPSAPMASTETLTPP	337		
Db	270	-SPKMSALGKKGRSALADSSGGVTLTL	SLKTFSPDRGCPTECTDSSGNSISTSPVHTTD	328		
Qy	338	ERNELVYASIEVQATQGNLSSA--ODY-----	RALYDTQAQNSDELDIASGIL	385		
Db	329	YTSVNGAALAAVASKERQVEDTEPPYDPFVDGRGP	IRALVDYVEVAEDELSPNSGDLF	388		
Qy	386	AVIL-EGEDGMVYERNQGQFVP	408			
Db	389	EKLEDEDEOGCKGRKDGVRGLP	412			

```

RESULT 10
T16613
hypothetical protein K10B3.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C:Accession: T16613
R:Gattung, S.
submitted to the EMBL Data Library, February 1996
A:Description: The sequence of C. elegans cosmid K10B3.
A:Reference number: Z18546
A:Accession: T16613
A:Status: preliminary; translated from GH/EMBL/DBDU
A:Molecule type: DNA
A:Residues: 1-2427 <GAT>
A:Cross-references: EMBL:U49941; NID:G1206038; PID:G1206048; PIDN:AAE53876.1; GSPDB:GN00
A:Experimental source: strain Bristol N2, clone K10B3
C:Genetics:
A:Gene: CESP.K10B3.10
A:Map position: X
A:Introns: 68/2; 96/3; 138/3; 665/3; 1052/3; 1104/3; 1173/2; 1995/2; 2187/3; 2235/2; 2313/3
C:Superfamily: spectrin alpha chain; calmodulin repeat homology; SH3 homology; spectrin/
C:Keywords: EF hand
F:990-1037/Domain: SH3 homology <SH3>
F:2050-2155/Domain: spectrin/dystrophin repeat homology <SPR>
F:1227-2309/Domain: calmodulin repeat homology <EFH>

```

Query Match	8.6%;	Score 185.5;	DB 2;	Length 2427;
Best Local Similarly	23.5%;	Pred. No. 0.0028;		
Matches 95;	Conservative 67;	Mismatches 149;	Indels 93;	Gaps 21;

QY	37	KUVEILLR---QRQAEBERYKKEIYQIARKAGGQYEMNSLRTSPFDSLKQGTENVGASHIOL	94
Db	705	EEVEAWLSEIEGGVASEYDKRLVSVONL---OKKIGLIESPNAHNDYVDGIK----	757
OY	95	ALALREELSLSEEFRR-----EROKEORRKYEAIMDRVOKSKLTYKTTESKAYDOKCR	149
Db	758	AOOFQEE-----EHFNAPVYIKQENLOOYRNALRDLPLEKRK-----RKLESHGQ-NQLFR	808
OY	150	DADDAQAEERVASNGHOKOVESQONKAKOC-----KESATEAERY-YRONIEOLER	200
Db	809	DIED-ELAWIR-----EKQYVASTNGSRDLIGVQNLTKKQALLAEIINHDSQIESVSS	862
OY	201	ARTWEQEHRTTCFAFOLQEF-DRLTILIRNALMWHCNOLSMQCVKDEDELYEEVRLTLEGC	259

```

Db      | : | | | | | | : : : | | |
863 AANDIQGCH-----FLAPETRDKLAQLRD-----NMRLISK--AEKRT--- 9010

QY      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
260 DVEGDINFIQSKSTGPREPAPVPYQNYNDREYTPLGSPIDQSCVIRKFSGLHGSP 3119

Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
902 DLDSDLOAH-YLSDANADA-----WMSEKEPIVGSTD-----YGKD 9388

QY      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
320 KTTPEAPAASTETILPPPERNELVYASIEVQATQ-----GNLSSAQDYRALYPTA 3717

Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
939 EDSAEALLKRRALLSLDLEAFKGTIEDLRKQASQCKYQEOPMGLGRDC--VIALDYOE 9966

QY      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
372 QNSDELIDISAGDILAVILEGDEGWMVTVERNQORQPFVSGYLEKL 415

Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
997 KSPREVSMMKGDVLTLLANARNDDMMKKEVNDROGEFVPAAYKRI 1040

```

```

RESULT 11
A:8580
echinococcal microtrichal antigen EMI3 - tapeworm (Echinococcus multilocularis)
C:Species: Echinococcus multilocularis
C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
C:Accession: A48580
R:Frosch, P.M.; Geier, C.; Kaup, F.J.; Muller, A.; Frosch, M.
Mol. Biochem. Parasitol. 58, 301-310, 1993
A:Title: Molecular cloning of an echinococcal microtrichal antigen immunoreactive in
A:Reference number: A48580; MUID:9324121
A:Accession: A48580
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-426 <FRO>
A:Cross-references: GB:M96565; NID:9304852; PID:9304853
A:Experimental source: larva
A:Note: sequence extracted from NCBI backbone (NCBIN:130067, NCBIP:130068)
C:Superfamily: SH3 homology
F:378-426/Domain: SH3 homology <SH3>

```

Query Match	8.28;	Score 177.5;	DB 2;	Length 426;
Best Local Similarity	21.28;	Pred. No. 0.001;		
Matches 97; Conservative	66;	Mismatches 168;	Indels 127;	Gaps 19;

```

0Y 42 LLRQAOAEHRYGKELVQIARKAGGOTFMNSLRPSF-----SLKOOTEVGSAH 92
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MIOERADIEKAYASNI-----RKFAARLEM-FLRGVEGATNTILSGIAEAEEDNABLS 55
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 93 QIALALREELR-----SLEEFR-----ROKORRKYEAIMDRPOKSLSYKTTMESK 142
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 56 NIAAGLIMPVQLGIKMWRERPFHKSISTSIKEKKNDFSEENAKOTWYKHKKVNNCK 115
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 143 AYDQCRDADDAEQAFFERYVSANGHOKOVESONKAKOKESATEAEHYVHONIEOLERAR 202
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 146 EYFHACK-----TVRSLOVOYQANKNEPFGPEQIRKIEDLKRIGIMEEKTR 163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 203 TEWEO-----EHRITCEAF-QIOEPRRLTILRNALVHNCQSMOCVKD--- 245
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 KATYEALSSLDVTPRYEDMT--QVEFKAAFERERK---LYEKDQALOMQVLDISA 217
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 246 ---DELVEEYRKLLEGCDVGIDINGFIQSKSTORREPPAPYONYVDREVTPLIGPSI 301
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 218 KPNLSQIFVGLRETVAKVDADADLKNMSLAVGYDMAPNEP-PEEY-----SPEM 266
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 302 QPSCCVYKRFSGGLHG-----PKTTBPAPAAST----- 330
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 --SALGKGGRALADGSSGGVTLTSLKTTSPDRGGPIPGTDSGNSISTSPVHTTAYGS 324
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 331 -----ETLIPPERNELVYASIEVOATQGNLSSA--ODY-----RALYDYTA 371
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 325 NSYDHGSGCAIRPSDYTSVNNAAAIASEKORVEDTTPPYDFVODGGRGYPIRALDYVG 384
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 372 QNSDELISAGDILAVIL--ESEDGMWYERNGOREYVP 408
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 385 VEADDELSPNSGDLFEKLEDEDEOQWCKRGKRGVGLP 422
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```


RESULT 15
MMAXIC
myosin heavy chain IC - Acanthamoeba castellanii
N:Contains: myosin ATPase (EC 3.6.1.32)
C:Species: Acanthamoeba castellanii
C:Date: 30-Sep-1990 #sequence_rev1sion 30-Sep-1990 #text_change 22-Jun-1999
A:Accession: A33891; C34448; A24146
R:Jung, G.; Korn, E.D.; Hammer III, J.A.
Proc. Natl. Acad. Sci. U.S.A. 84, 6720-6724, 1987
A:Title: The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-I-like and non-my
A:Reference number: A33891; MUID:88016163
A:Accession: A33891
A:Molecule type: DNA
A:Residues: 1-1168 <JUN>
A:Cross-references: GB:J02974; NID:g155624; PIDN:AAA27707.1; PID:g155625
A:Note: this gene and protein are called MIB in this paper
R:Bireska, R.; Lynch, T.J.; Martin, B.; Korn, E.D.
J. Biol. Chem. 264, 19340-19348, 1989
A:Title: The localization and sequence of the phosphorylation sites of Acanthamoeba myos
A:Reference number: A34448; MUID:90037074
A:Accession: C34448
A:Molecule type: protein

QY	100	ELKSLSEEFRRER-----	-----OKEQRKTY-----	-----EAIIDRQKSLST----	133
		: :	:	: : :	
Db	656	ESVSLSEELDRIVFSTANKIQRFLRKTAMRKTYEYEVKGGNDALVANKKERRLSLERPF			715
QY	134	-----YKKTWESKRAYDQK-----	-----CRDADA-----	-----EOAFERYSANGHOKOVE	171
Db	716	KTDYINRYONFKTECDIGDKGTEKVFELFADLCNNLDKSFMSKVERRIIMVLTSMN-MFLVA			774
QY	172	KSOKKAAQCQKESAFEARVRYRONIEQLERKRTMEQEHRTTCFAFO-----			217
Db	775	IDPKKDK-----	-----IKKKKPPLYLVKLR-RIDFNKIGSTITSLPDNDFMLISVNGEHSN		825
QY	218	LOEPRDLTIIRNALMFWHCNOLSMQCVKDELYEVRLLTEG----	CDV-----	EGDI	265
		: : : : :	: : :	:	
Db	826	LLECRRTLELGLVLLKKNIPSVRIQ-----	FATFNVTLLKGCKTYCVKRTIDRPQSGDKAV		879
QY	266	NGFIQSKSTGCEPPAPVP-----	-----YONYNDREVTPLIGS-----		298
Db	880	KGTGVSAVAPGL-PPSSAPNIOAPQETSIGSAPTVAEQSYKQD-----	ILAKGGGGGGGGRG		934
QY	299	---PSIQPCGCVIKRFSG-GLLHSGP-----	KTPPSAPAASTEHLPTLPBRLNELVYASIEIV		349
		: : : : :	: : : :	:	
Db	935	RGGRS--PSCGVSVSRPSPGGGGGGGPFSPFGGRPSPSCPPAAS--	ADPGD-----		980
QY	350	QATGQNLNSSAQDYRALDYDTAQNDELDISAGDILAVILLEGEDGMMVTERNGQGRFPVG			409
		: : : : :	: : : :	:	
Db	981	-----ARALYDDEAENDELTFNEGAVVTVINKSPDMWEGELNGQRGVFA			1022
QY	410	STYLE	413		
		: :			
Db	1028	SYVE	1031		

Search completed: September 15, 2000, 11:59:33
Job time: 5715 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2000, 11:07:13 ; Search time 43.01 Seconds

(without alignments)
299.011 Million cell updates/sec

Title: US-09-068-377-1

Perfect score: 2152

Sequence: 1 MMAQLQFRDAFCRCDFTAHT.....MTVERNGRGFGPESYLEKL 415

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	346	16.1	670	1 YBY2_SCHPO	Q10199 schizosacch
2	345.5	16.1	927	1 CC15_SCHPO	Q09822 schizosacch
3	176.5	8.2	1113	1 MYSD_DICDI	P34109 dictyostell
4	171	7.9	1168	1 MYSC_ACACA	P10569 acanthamoeb
5	167	7.8	546	1 SRC8_MOUSE	Q60598 mus musculu
6	166.5	7.7	633	1 YHR4_YEAST	P38822 saccharomyc
7	166	7.7	550	1 SRC8_HUMAN	Q14247 homo sapien
8	165	7.7	526	1 NCF2_HUMAN	P13878 homo sapien
9	163.5	7.6	2477	1 SPCN_CHICK	P07751 gallus gall
10	162	7.5	527	1 NCF2_BOVIN	O77775 bos taurus
11	161	7.5	2472	1 SPCN_HUMAN	O13813 homo sapien
12	160.5	7.5	642	1 YB65_SCHPO	Q09496 schizosacch
13	160.5	7.5	777	1 VA21_SCHPO	Q09697 schizosacch
14	159.5	7.4	482	1 R167_YEAST	P39743 saccharomyc
15	158	7.3	563	1 SRC8_MOUSE	Q01406 gallus gall
16	158	7.3	1920	1 PCNT_MOUSE	P48725 mus musculu
17	151.5	7.0	2415	1 SPCA_DROME	P13395 drosophila
18	149.5	6.9	520	1 ITSN_HUMAN	O15811 homo sapien
19	147.5	6.9	392	1 NCF1_BOVIN	O77724 bos taurus
20	145.5	6.8	1935	1 MYSB_PTG	P73293 sus scrofa
21	144.5	6.7	1935	1 MYSB_HUMAN	P12883 homo sapien
22	144	6.7	261	1 LAS1_HUMAN	Q14847 homo sapien
23	144	6.7	263	1 LAS1_MOUSE	O61792 mus musculu
24	144	6.6	695	1 XE7_HUMAN	Q02040 homo sapien
25	141	6.6	1679	1 Y109_YEAST	P40457 saccharomyc
26	138.5	6.4	2663	1 CENE_HUMAN	Q02224 homo sapien
27	138	6.4	1937	1 MYSP_HUMAN	P15353 homo sapien
28	138	6.4	2349	1 TPR_HUMAN	P12270 homo sapien
29	137.5	6.4	390	1 NCF1_HUMAN	P14598 homo sapien
30	137.5	6.4	1934	1 MYSB_MESAU	P13540 mesocricetu
31	137.5	6.4	1935	1 MYSB_RAT	P02564 rattus norv
32	137	6.4	2649	1 BRAL_HUMAN	Q03001 homo sapien
33	136.5	6.3	2418	1 SPCA_HUMAN	P02549 homo sapien

34	135.5	6.3	390	1 NCF1_MOUSE	Q09014 mus musculu
35	135.5	6.3	744	1 YEF7_YEAST	P43556 saccharomyc
36	135.5	6.3	845	1 VAV_MOUSE	P27870 mus musculu
37	135.5	6.3	1391	1 MST2_DROXY	Q08696 drosophila
38	134.5	6.2	1395	1 SP41_YEAST	P38904 saccharomyc
39	134	6.2	486	1 HSI_MOUSE	P49710 mus musculu
40	134	6.2	1147	1 MYSB_ACACA	P19706 acanthamoeb
41	134	6.2	1427	1 MYST_HUMAN	P10622 homo sapien
42	133.5	6.2	1181	1 MYSC_DICDI	P42522 dictyostell
43	133	6.2	1085	1 YAPF_SCHPO	Q09863 schizosacch
44	133	6.2	1790	1 USO1_YEAST	P25386 saccharomyc
45	132.5	6.2	878	1 VAV2_HUMAN	P52735 homo sapien

ALIGNMENTS

RESULT 1	
YBY2_SCHPO	
ID YBY2_SCHPO	STANDARD: PRT; 670 AA.
AC Q10199;	
DT 01-OCT-1996 (Rel. 34, Created)	
DT 01-OCT-1996 (Rel. 34, Last sequence update)	
DT 15-JUL-1999 (Rel. 38, Last annotation update)	
DE HYPOTHETICAL 75.2 KDA PROTEIN C11C11.02 IN CHROMOSOME II.	
GN SPBC11C11.02.	
OS Schizosaccharomyces pombe (fission yeast).	
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;	
OC Schizosaccharomycetaceae; Schizosaccharomycetes.	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=972;	
RA Wood V., Rajadream M.A., Barrell B.G., Beck A., Reinhardt R.;	
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.	
CC -1- SIMILARITY: SOME, TO S.POMBE CDC15 AND TO YEAST YMR032W.	
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.	
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (See http://www.isb.ch/announce/	
CC or send an email to license@isb-sib.ch).	
CC -----	
DR EMBL: AL031528; CAA20684.1; -.	
DR HSSP: P19174; IHSQ.	
DR PFAM: PF00611; FCH; 1.	
DR PFAM: PF00018; SH3; 1.	
DR PRINTS: PR00452; SH3DOMAIN.	
DR PROSITE: PS50002; SH3; 1.	
KW Hypothetical protein; SH3 domain.	
FT DOMAIN 607..670	
SQ SEQUENCE 670 AA; 75167 MW; 03551FBB56FC91AE CRC64;	

Query Match 16.1%; Score 346; DB 1; Length 670;

Best Local Similarity 18.6%; Pred. No. 9.2e-14; Index 264; Gaps 13;

Matches 133; Conservative 84; Mismatches 191; Index 264; Gaps 13;	
QY 7 FRDAFCRCDFTAHTGYEVLLRLDLGRKCKDVEELLDRQRAQAEERKGLVQIARRAGG 66	
DB 18 FSNYFPGCAN---DEGYHALLSRSDYKHNIEELRSFYHERANIEEDYAKMAKISRTTFS 74	
QY 67 QTENSLRFSFDSLKOOTEVNGSAHIQIALALDEL-RSLDEPREROKERKKYEALIMDR 125	
DB 75 SLETGCLIKESVQYKAEVDMAKSHLQISQLDDVENATFRYAASLKKKKMIVSGIEK 134	
QY 126 VQSKSLVKKKTKESKRAYQKCRDADDAQAEFRYSANG--HQKQVKSQNKAKQCKES 183	
DB 135 VHKDKSLKADKADKADKADKADKADKADKADKADKADKADKADKADKADKADKADKAD 188	
QY 184 ATEAEVRYRONIBOLERRARTWEDEHRTTCFAFOLQEFDRLLTIRNALWVHCNQLSMQCV 243	

```

Db 189 ITASSSDYQSAVAAARSDYANRMTNEMRSTCKLQDIEERHFLKSWMTTLTLLSRSCF 248
QY 244 KDELVEEVRLLTLEGGDEVDINGFIQSKSTGREPPAPVQYQNYDREV----- 292
Db 249 NDDQACERIRKRLNLCGCSVSQDYLEFIDAKSTGTGTPQPKRYDYKKGVPDSDVLYQAN 308
QY 293 ----- 292
Db 309 FORAQTKIENDNMLNRPYVLSAATARNESFENLPTPSAIOSTITVSSNSSONGRSSP 368
QY 293 -----TPILG-----SPSIOPSCGVIKRFGSLHGSFK--TT 322
Db 369 KKSFLSKFKLTSRPSTPVNGTAPDALSSPRNDSPLTSADEQMKHLS--LQEEPKQNP 426
QY 323 PSAPAASTETLTPPERNEL----- 342
Db 427 PAAGGAFNNSMTLPPRYNELGSLSPNSVSTFEDSRPVNTPSRROQIQEEFGSVLQMEN 486
QY 343 -----VYAS-----IEVQATQGNL----- 356
Db 487 RAVSPVYDSRRKNGSSFTLKRSPKRPSSSLSONASRLPRSLTPGNLEPNYDFGVYVD 546
QY 357 -----NSSAQD-- 362
Db 547 PASGAPTDEPYTDROSSFVDDTINTKAGTNSRLSLPAYPTDGGDTSIDNPSTDGQ 606
QY 363 -----YRALYDYTONSDELDISAGDILAVILEGDEGWM-----TVERNGRGFPVGYL 413
Db 607 RILGYVSLYDYDAIPEIISFRKGDITAVLYKLYEDGWEGFVYGGEDHNGQPSNRYR 666
QY 414 KL 415
Db 667 EI 668

```

RESULT 2

CC15_SCHPO STANDARD; PRT; 927 AA.

```

AC 009822; 014365;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE CELL DIVISION CONTROL PROTEIN 15.
GN CDC15 OR SPAC2068.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetaceae; Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RX MEDLINE; 95360987.
RA Fankhauser C., Raymond A., Cerutti L., Utzig S., Hofmann K.,
RA Simanis V.;
RT "the S. pombe cdc15 gene is a key element in the reorganization of F-
actin at mitosis."
RL Cell 82:435-444(1995).
RN [2]
RP REVISIONS TO N-TERMINUS.
RA Fankhauser C., Raymond A., Cerutti L., Utzig S., Hofmann K.,
RA Simanis V.;
RN Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RL [3]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA Badcock K., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;
RA Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: AFTER THE ONSET OF MITOSIS, FORMS A RING-LIKE STRUCTURE
WHICH CO-LOCALIZES WITH THE MEDIAL ACTIN RING. APPEARS TO MEDIATE
CYTOSKELETAL REARRANGEMENTS REQUIRED FOR CYTOKINESIS. ESSENTIAL
FOR VIABILITY.
CC -1- DEVELOPMENTAL STAGE: PEAKS IN EARLY MITOSIS BEFORE SEPARATION.
CC -1- DOMAIN: THE N-TERMINAL REGION IS IN A COILED COIL STRUCTURE.

```

```

CC -1- PTM: PHOSPHORYLATED.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: SOME, TO S.POMBE SPB1C11.02.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X66179; CA60115.1; -.
DR EMBL; Z95334; CAB08599.2; -.
DR HSSP; P07751; ITUD.
DR PFAM; PF00611; FCH; 1.
DR PFAM; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PROSITE; PS00002; SH3; 1.
KW Mitosis; Cytoskeleton; SH3 domain; Phosphorylation; Coiled coil.
FT DOMAIN 24 110 FCH.
FT DOMAIN 108 207 COILED COIL (POTENTIAL).
FT DOMAIN 866 927 SH3.
SQ SEQUENCE 927 AA; 102119 MW; FDCFE7E0AAA3D247D CRC64;

```

Query Match

Best local similarity 25.9%; Score 345.5; DB 1; Length 927;
Matches 96; Conservative 74; Mismatches 169; Indels 31; Gaps 8;

```

QY 3 AQOLFRAFWCRDFTAHGTGEVLLQLRLDGRKCKDVEILLRORQAQEEYKGEVQIAR 62
Db 18 SSVKFRDNFMGSE--DAGMDALMSRTKSSLSVLESIDEFYAKRASIEREYASKLOELAA 74
QY 63 KAGGQTEFMSNLSRTSFDSLKQOOTENVGSAHQLALALREELRS--LEEFEROKRQKYYEA 121
Db 75 SSADIPVEGSLTNILISMKRTETGSMAKAHEVSOQITELRNKRTIITDQKVVAAAN 134
QY 122 IMDRVQSKSLVYKTKMESKRAYDQCRDADDAQAFERYASANGHOKOVERSONKAKOCK 181
Db 135 AIELYQKKTALIEIDLEKKDAYEYSCNKLNSYVRQKKMG-----RELKYNLAKINQAA 190
QY 182 ESATAEVRYRQNTQELRARTWEQEHRTTCEAFQLOEFDRLTILNALMWHCNLSMQ 241
Db 191 LAVKMDAEYRETNELLTYTRWIDRWTCYAFQHIIEEYRLFELKTNMMAVANIISTA 250
QY 242 CVKDELVEEVRLLTLEGGDEVDINGFIQSKSTGREPPAPVQYQNYDREV-----PTPL 295
Db 251 CVKDESCERILTLTENTNIDEDITQMIQNDGTTIP-PLPERNDYFKENGLYNDIDOL 309
QY 296 IGSPSIOPSCGVIKRFGSLHGSFKTTPSAPAASTETLT-----PT-----PERN 340
Db 310 ISKAPSYPRSSRSASASLASSPTRSAPRPKTS-ELVSEVYVSPPTSLHSPVKVSN 368
QY 341 ELVYASIEVQ 350
Db 369 EOYEQVTEVE 378

```

RESULT 3

MYSD_DICDI STANDARD; PRT; 1113 AA.

```

AC P34109;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MYOSIN ID HEAVY CHAIN.
GN MYOD OR DMID.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Dictyostellida; Dictyostellum.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 604-610; 733-742 AND 914-928.
RC STRAIN-AX3;

```


RX MEDLINE; 93315475.
 RA Jung G., Fukui Y., Martin B., Hammer J.A. III;
 RT "Sequence, expression pattern, intracellular localization, and
 RT targeted disruption of the Dictyostellium myosin ID heavy chain
 isoform."
 RL J. Biol. Chem. 268:14981-14990(1993).
 CC -I- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
 CC ACTIVITY THAT IS ACTIVATED BY ACTIN. MYOSIN ID MAY HAVE A ROLE IN
 CC CHEMOTAXIS AND AGGREGATION; IT COULD SERVE TO STABILIZE AND EVEN
 CC RETRACT CORTICAL STRUCTURES, SUCH AS PSEPODPODS AND LAMELOPODS.
 CC -I- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
 CC -I- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION JUST BENEATH THE
 CC PLASMA MEMBRANE IN THE ANTERIOR PSEUDOPOD AT THE LEADING EDGE OF
 CC THE CELL.
 CC -I- DOMAIN: TH.1 BINDS DIRECTLY TO ANIONIC PHOSPHOLIPID MEMBRANES;
 CC MYOSIN I COULD THEREFORE MOVE ACTIN RELATIVE TO MEMBRANES AND
 CC VICE VERSA. TH.2 AND SH3 BIND TIGHTLY TO F-ACTIN; THIS TOGETHER
 CC WITH THE NUCLEOTIDE-SENSITIVE SITE IN THE HEAD, ALLOWS SINGLE
 CC MOLECULES OF MYOSIN I TO CROSS-LINK ACTIN FILAMENTS.
 CC -I- SIMILARITY: BELONGS TO THE MYOSIN TYPE I FAMILY (SMALL MYOSINS).
 CC -I- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L16509; -; NOT_ANNOTATED_CDS.
 CC DR PIR; A47106; A47106.
 CC DR HSSP; P29354; 1GRI.
 CC DR DICTYDB; DD01048; MYOD.
 CC DR PFAM; PF000018; SH3; 1.
 CC DR PFAM; PF00063; myosin_head; 1.
 CC DR PRINTS; PRO0193; MYOSINHEAVY.
 CC DR PRINTS; PRO0452; SH3DOMAIN.
 CC DR PROSITE; PS50002; SH3; 1.
 CC KW Myosin; Actin-binding; ATP-binding; Chemotaxis; SH3 domain;
 CC Multiene family.
 CC FT DOMAIN 1 680 GLOBULAR HEAD-LIKE DOMAIN.
 CC FT NP_BIND 1113 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.
 CC FT NP_BIND 101 ATP (BY SIMILARITY).
 CC FT DOMAIN 681 956 TAIL HOMOLOG REGION 1 (TH.1).
 CC FT DOMAIN 957 966 GLY/PRO/ALA-RICH (TH.2).
 CC FT DOMAIN 1022 1110 GLY/PRO/ALA-RICH (TH.2).
 CC FT DOMAIN 962 1021 SH3.
 CC FT DOMAIN 1016 1113 ALA/GLY/PRO-RICH.
 CC SQ SEQUENCE 1113 AA; 124637 MW; 07F6B6C7FA7C2A08 CRC64;

Query Match 8.2%; Score 176.5; DB 1; Length 1113;
 Best Local Similarity 21.9%; Pred. No. 0.002;
 Matches 89; Conservative 65; Mismatches 148; Indels 105; Gaps 20;

OY 71 NSLRTSFDLSKQOTENVGSAHQLALALRELSLEEFREKQ-EQKKYEAIIMDR----- 125
 DB 652 SCVETILKMDLEPKQYKCKTKIFIRAPETVNLBELRKRYFTYKNIQRPFLRTIM 711
 OY 126 -----VQSKSLSYKRTMSKAYDOKCRDAD--DAEQAER--VSANGHOK-----OVE 171
 DB 712 SYYSIOKGAADSMKSNKERRRLSIEPYOGDIYINRENELDIYKKNENKIMPTHAV 771
 OY 172 KSONKAKCKES-----ATEAEKVYKONIEQLERATEW--EQEHR-----T 211
 DB 772 NKYDRSKORRVLILSDTAIYPIARE-----KNKDKERKRPRWYVYQKRRLLAGIT 825
 OY 212 TCEAFLOEEDRLTILRNLMVHCNOLSMQCVKDELEYEVRILTCEGDEVDING-FIQ 270
 DB 826 SVELSKLS--DGEVVLK-----RANEDHDLFE-----CRKTELEGLTIK 863

OY 271 SKSTGREPPAPVRYQNYDREVT-----LIGSPSIQPSGCVIKRPSGLHGGSP 319
 DB 864 AYKKG---TLKRNVNSIGVAIKASKOGKGERITLPEKGIPESEYFK-----GTK 913
 OY 320 KTTFS--APASTETLTPTPERNELVYASIEVQATQNLNSAOD-----YRAL 366
 DB 914 VSPFSDGLPADTVPLTLP--PE--SLPVVSIPIYKPAKNAKNAPOMSGPASNVPKSAAL 970
 OY 367 YDYAONSELDISAGDILAVILEGEDMWYTERRGQFVGYSTLE 413
 DB 971 YDFAESSEMEISFEKGDILITVLDQSSGDMDAELKGRGKVPSTNLTQ 1017

RESULT 4
 MYSC_ACACA STANDARD; PRT; 1168 AA.
 AC P10569;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE MYOSIN IC HEAVY CHAIN.
 GN MIT.
 OS Acanthamoeba castellanii (Amoeba).
 OC Eukaryota; Acanthamoebidae; Acanthamoeba.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 88016163.
 RA Jung G., Korn E.D., Hammer J.A. III;
 RT "The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like
 RT and non-myosin-like sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 84:6720-6724(1987).
 RN [2]
 RP PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE; 86259656.
 RA Hammer J.A. III, Jung G., Korn E.D.;
 RT "Genetic evidence that Acanthamoeba myosin I is a true myosin."
 RL Proc. Natl. Acad. Sci. U.S.A. 83:4655-4659(1986).
 RN [3]
 RP PHOSPHORYLATION SITE.
 RX MEDLINE; 90037074.
 RA Brzeska H., Lynch T.J., Martin B., Korn E.D.;
 RT "The localization and sequence of the phosphorylation sites of
 RT Acanthamoeba myosins I. An improved method for locating the
 RT phosphorylated amino acid."
 RL J. Biol. Chem. 264:19340-19348(1989).
 CC -I- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
 CC ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
 CC -I- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
 CC AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
 CC -I- DOMAIN: TH.1 BINDS DIRECTLY TO ANIONIC PHOSPHOLIPID MEMBRANES;
 CC MYOSIN I CAN THEREFORE MOVE ACTIN RELATIVE TO MEMBRANES AND VICE
 CC VERSA. TH.2 AND SH3 BIND TIGHTLY TO F-ACTIN; THIS TOGETHER WITH
 CC THE NUCLEOTIDE-SENSITIVE SITE IN THE HEAD, ALLOWS SINGLE MOLECULES
 CC OF MYOSIN I TO CROSS-LINK ACTIN FILAMENTS.
 CC -I- MISCELLANEOUS: THIS ORGANISM EXPRESSES AT LEAST THREE ISOFORMS OF
 CC MYOSIN I HEAVY-CHAIN, ENCODED BY GENES MIA, MIB, AND MIC.
 CC -I- SIMILARITY: THE GLOBULAR HEAD DOMAIN IS SIMILAR WITH THE GLOBULAR
 CC HEAD SEQUENCE OF MUSCLE MYOSINS. BY CONTRAST, THE COILED-COIL ROD-
 CC LIKE REGION FOUND IN MANY MYOSIN HEAVY CHAINS IS REPLACED BY A
 CC CARBOXYL-TERMINAL DOMAIN CONTAINING A REGION RICH IN GLYCINE,
 CC ALANINE, AND PROLINE.
 CC -I- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -I- CAUTION: WAS ORIGINALLY THOUGHT TO BE MYOSIN IB.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; J02974; AAA27707.1; -.

DR PIR: A33891; MMXIC.
 DR HSP: P08799; 1MD.
 DR PFAM: PF00018; SH3: 1.
 DR PFAM: PF00063; myosin_head.1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR PROSITE: PS0002; SH3: 1.
 KM Myosin; ATP-binding; Phosphorylation; Multigene family; SH3 domain.
 FT DOMAIN 1 670
 FT NP_BIND 671 1168
 FT NP_BIND 101 108
 FT DOMAIN 671 922
 FT DOMAIN 923 978
 FT DOMAIN 1034 1168
 FT DOMAIN 976 1035
 FT MOD_RES 311 311
 SO SEQUENCE 1168 AA; 127309 MW; D07084B373A37A32 CRC64;

Query Match Best Local Similarity 7.9%; Score 171; DB 1; Length 1168;
 Matches 97; Conservative 53; Mismatches 116; Indels 158; Gaps 21;

QY 100 EELRSLSEFRER-----OKEORRY-----EALMDRYOKSKLSL--- 133
 DB 656 ESYFSLSELRDRTFVSYANKIQRLFKTKMKRYEYVKKGNDALVNNKERRRLSLRPF 715
 QY 134 -----YKTKMESKRAYDOK-----CRDADA-----EQAFRVSANGHOKOVE 171
 DB 716 KIDYITRQNFKDKDIGKTEKYLFDLCLNNLDSFWSKVERIRIWLTSMA-MELVA 774
 QY 172 KSONKAKOCKESATEAEERYRONIEQLERARTEMEQEHRTCEAFQ----- 217
 DB 775 IDPNKDK-----IEKKVKPFLVYLKR-RIDPNKISITLSPLQDNFMLSIVNGHSN 825
 QY 218 IOEPRRLTLIRNALWVHCNOISMOCVKDELYEEVRLTLES---CDV-----EGDI 265
 DB 826 LLECRKTELGLVLTKNHPVRIQ-----FADTFNVLTKGKTCVAKFLRDQGGGKGV 879
 QY 266 NGFIOSKSTGRRPAPVP-----YONYVDREVTPLIGS----- 298
 DB 880 KGTQSVAPGL-PPSSAPNIIQAPOETSGGASTYAEOSYDQ-----TLGKGGGGGGGRG 934
 QY 299 ---PSIOPSCVIRKFS-GLIHGSP-----KTPSAPASTETLTPPEENELVYASIEV 349
 DB 935 RGGFS--PSGAVSPRSPGGGGSPFGGRSPSPGPAAS--AGCPG----- 980
 QY 350 QATGNNLSSADYRALDYTAONDELDSAGDILAVILLEGDGWVYERNRGGEVPG 409
 DB 981 -----ARALYDFAENPDELTFNEGAVTVINKSNPDWEGELNGORGVFPA 1027
 QY 410 STYLE 413
 DB 1028 SYVE 1031

RESULT 5
 SRC8_MOUSE STANDARD; PRT; 546 AA.
 ID SRC8_MOUSE
 AC 060598;
 DT 01-NOV-1997 (rel. 35, Created)
 DT 01-NOV-1997 (rel. 35, Last sequence update)
 DT 15-DEC-1998 (rel. 37, Last annotation update)
 DE SRC SUBSTRATE CONTACTIN.
 GN CTIN OR EMS1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C;
 RX MEDLINE; 94268839.
 RA Migliarese M.R., Mannion-Henderson J., Wu H., Parsons J.T.,

RA Bender T.P.;
 RT "The protein tyrosine kinase substrate cortactin is differentially
 RT expressed in murine B lymphoid tumors.";
 RL Oncogene 9:1989-1997(1994).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 125-138; 273-289 AND 534-543.
 RX MEDLINE; 94033284.
 RA Zhan X., Hu X., Hampton B., Burgess W.H., Friesel R., Maciag T.;
 RT "Murine cortactin is phosphorylated in response to fibroblast growth
 RT factor-1 on tyrosine residues late in the G1 phase of the BALB/c 3T3
 RT cell cycle.";
 RL J. Biol. Chem. 268:24427-24431(1993).
 CC -1- FUNCTION: MAY CONTRIBUTE TO THE ORGANIZATION OF CELL STRUCTURE.
 CC THE SH3 MOTIF MAY FUNCTION AS A BINDING REGION TO CYTOSKELETON.
 CC CELLULAR GROWTH REGULATION IN TRANSFORMED CELLS MAY CONTRIBUTE TO
 CC DETECTED IN B LYMPHOCYTES OR PLASMA CELLS.
 CC -1- SIMILARITY: TO HEMATOPOIETIC LINEAGE CELL SPECIFIC PROTEIN.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC or send an email to license@sib.ch).
 CC EMBL; U03184; AAA19689.1; -.
 DR HSP: P19174; 1HSQ.
 DR HSP: MGI:99695; CTIN.
 DR PFAM: PF00018; SH3: 1.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR PROSITE: PS0002; SH3: 1.
 KW Phosphorylation; Repeat; SH3 domain; Cytoskeleton.
 FT DOMAIN 83 329
 FT REPEAT 83 119
 FT REPEAT 120 136
 FT REPEAT 157 193
 FT REPEAT 194 230
 FT REPEAT 231 267
 FT REPEAT 268 304
 FT REPEAT 305 329
 FT DOMAIN 488 546
 FT CONFLICT 9
 SO SEQUENCE 546 AA; 61260 MW; 8F93A026AD1D6D4F CRC64;

Query Match Best Local Similarity 7.8%; Score 167; DB 1; Length 546;
 Matches 92; Conservative 41; Mismatches 135; Indels 128; Gaps 15;

QY 40 EELLRQRAQAEERYKGLVQIRKAGGQTEKNSLRTSPDSLKQOTENGSAHIDALALR 99
 DB 254 EKLQHSQKQDYKTG-----FGKFGVQSE-----RQSSANGFY-----K 290
 QY 100 EELRSLSEFRERQKQKRYEALIMRVOKSKLSYKTKMESKRAYDOKCRDADAQAF 159
 DB 291 ERLAKHEPOODYAKGFGKGYGQKRMOKN-ASTFEVYQVPSAY----- 334
 QY 160 RVSANGHOKOVEKSONKAKOCKESATEAEERYRONIEQLERARTEMEQEHRTCEAFQ 219
 DB 335 -----QATVP-----IEAVTSKTSNIRANFENLAKER---RQDRRAAEARQ 375
 QY 220 EFDRLTLIRNALWVHCNOISMOCVKDELYEEVRLTLEGDEVDINGFTQSKSTGREPP 279
 DB 376 -----RAKEROQDEARAKLE-----QARAKQTPP 403
 QY 280 APVYQNYDR-----EVTPLIGSPSIOPSCGVIRKFSGLIHGSPKTPSPAPAA--- 328
 DB 404 ASPSPQIEDRPPSSPIYEDAAPKAPESYRGS-----EPEPEISIEAAGIP 450

OY 329 ---STETLTPEERNEELVASTIEV---OATGONLSSADYR---ALYDTAONSDEL 377
 Db 451 EAGSOGGLTYT---SEPVYETTEAPGHYQAEEDDYDESDLGITAIALYDYQAADDEI 507
 OY 378 DISAGDILAVILEGEGDGMWTVERNRGORGFVPGSYLE 413
 Db 508 SPDPDITITNEMIDGMWRCVKGRYGLFPANYVE 543

RESULT 6

YHR4_YEAST

ID YHR4_YEAST STANDARD; PRT: 633 AA.

AC P38822;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE HYPOTHEETICAL 71.2 KDA PROTEIN IN ERP5-ORC6 INTERGENIC REGION.
 GN YHR14W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 NC Saccharomycetaceae; Saccharomyces.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RX MEDLINE: 94378003.

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
 RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
 RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
 RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
 RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
 RA Vaudin M.;

RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 VII.";

RL Science 265:2077-2082(1994).

CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.

CC -1- SIMILARITY: TO S.POMBE SPBC12C2.05C.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC -----
 CC EMBL: U00059; AAB6850.1; -.

DR PIR: S48956; S48956.

DR PFAM: PF00611; FCH; 1.

DR PROSITE: PS0018; SH3; 2.

DR PROSITE: PS50002; SH3; 2.

FW Hypothetical protein: SH3 domain.

FW DOMAIN 493 555 SH3.

FW DOMAIN 577 633 SH3.

FW SEQUENCE 633 AA; 71171 MW; 5C73DAC69611B41 CRC64;

Query Match 7.7%; Score 166.5; DB 1: Length 633;
 Best Local Similarity 16.3%; Pred. No. 0.0041;
 Matches 102; Conservative 81; Mismatches 175; Indels 267; Gaps 22;

OY 34 KWCKVEEILRORAAEERYGKELVQIARAGQTEMSLRIS----- 76
 Db 27 KWLKIDEGFYRRARAKLEKYSERLSLAIFYNKKSTSVPLISVGTPTTSGSIAAGV 86

OY 77 --FDSLKOOTEVNSAHIQALALREEL-RSLDEPEREKORRKYEAIMDRYQSKLSL 133
 Db 87 VAWNRIISQTDWISKDHDQLSTDFENHVANQJSLFTKIDMTLSKNGCFNMVMNKKDNI 146

OY 134 YKKTESKRAYOKCRDADDAEAFERVSANGHOKOVEKSONKAKCKCSATFAEAVYNG 193
 Db 147 YHELEKAKKDIDEACSTMEMANRRYTKASNDNRKKKLDEKEMKCKNE-----YLI 199

OY 194 NIOLEFARTEWDEHRTTCEAFQ-----LOEFDRL-TILRNALWVHC---NOLSM 240
 Db 200 KINQANTRTKDY-----YFQDVEVLDLQDVNEAKTLFLNDLMLKAASYENDLGA 250
 OY 241 QCYK-----DDELY----- 249
 Db 251 NVSKRIQAANSVYKONKPSLNTALIFKHLNKNMKEPODFYKPSPVWHDDEKFAVPSLE 310
 OY 250 -EEVRLTEGCDVEGDINGFIQSKSTG-----REPPAPVPYQNYD--- 289
 Db 311 VEDLRITL- AKANDYNS-LQDKTONELSKLTLNKIKHEKTNEDINATKRYDLKE 367
 OY 290 --REVPLIGPSI-----QPSGVYKRF----- 311
 Db 368 YLWVSPFTSHETLKQAEVQIESIQNNVPEEYDLSTDNDLSKTKKSGIFSFKINIL 427
 OY 312 -----SGLH-----GSKTTPSAPASTETLP 335
 Db 428 NVDSKPPSSGSGTNGNGNGLPHTSLFNTSRRTLGSAAPNAGEDSDNNSIRTTSTNTRK 487
 OY 336 TPE-----RNELYYASI-----EVOATQGN----- 355
 Db 488 TTQNSSPDGKKNKYLAYVQADDEITITTPGDKISLVARDGSGWTKINNDTGETGLVPT 547
 OY 356 -----LNSSA-----QDYRALDYTAQNSDELDSAGDILAVI 388
 Db 548 TYIRISSAAVYKANDRGAPPEPRRSTLPYRTEMAIYAEAGDDEISIDPDITIVI 607
 OY 389 LEGED--GMWTVERNRGORGFVPGSY 411
 Db 608 -RQDDGSGWTVYGRDGLKGLFPITSY 631

RESULT 7

SRC8_HUMAN

ID SRC8_HUMAN STANDARD; PRT: 550 AA.

AC Q14247;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE SRC SUBSTRATE CONTRACTIN (AMPLAXIN) (EMSI ONCOGENE).
 GN CTN OR EMSI.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

NC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE-BREAST;
 RX MEDLINE: 92195667.

RA Schuring E.M.D., Verhoeven E., Mool W.J., Michalides R.J.A.;

RT "Identification and cloning of two overexpressed genes, UZ1B31/PRAD1
 RT and EMS1, within the amplified chromosome 11q13 region in human
 RT carcinomas.";

RL Oncogene 7:355-361(1992).

RL [2]

RP SEQUENCE FROM N.A.

RC TISSUE-BREAST;

RX MEDLINE: 92333650.

RA Schuring E.M.D., Verhoeven E., Litvinov S., Michalides R.J.A.;

RT "The product of the EMS1 gene, amplified and overexpressed in human
 RT cell-substratum contact sites.";

RL Mol. Cell. Biol. 13:2891-2898(1993).

CC -1- FUNCTION: MAY CONTRIBUTE TO THE ORGANIZATION OF CELL STRUCTURE.
 CC THE SH3 MOTIF MAY FUNCTION AS A BINDING REGION TO CYTOSKELETON.

CC TYROSINE PHOSPHORYLATION IN TRANSFORMED CELLS MAY CONTRIBUTE TO
 CC CELLULAR GROWTH REGULATION AND TRANSFORMATION.

CC -1- SIMILARITY: TO HEMATOPOIETIC LINEAGE CELL SPECIFIC PROTEIN.

CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its


```

QY 289 -DREYTPPL-----ISPSIOPSCGVYIKRSGILHSGPKTPPSAPASTETLP-TEBRNE 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 NEROVAOLKKKDYKATVVASVVDQDSFGF-----APLOPQAAPPPRRKTPB--- 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 LVYASLEVOATGNNSSNADQALDYTAONSDELISAGDILAVILLEGEDGMWVERN 401
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 236 -IFRALEGEA-----HRLVLFGEVPEKKEELQVMPGNLIVLKKGNMWTAVMEN 283
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 402 GORGEVPGSYLE 413
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 GOKGLVPCNYLE 295
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
SPCN_CHICK STANDARD: PRT: 2477 AA.
AC P07751:
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SPECTRIN ALPHA CHAIN, BRAIN (SPECTRIN, NON-ERYTHROID ALPHA CHAIN)
DE (FOODRIN ALPHA CHAIN) (SPTANI).
CN SPTA2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
CN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE: 89093238.
RA Wasenius V.-M., Saraste M., Salven P., Eraama M., Holm M.,
RA Lehto V.-P.;
RT "Primary structure of the brain alpha-spectrin.";
RL J. Cell Biol. 108:79-93(1989).
RN [2]
RP REVISIONS.
RA Wasenius V.-M., Saraste M., Salven P., Eraama M., Holm L.,
RA Lehto V.-P.;
RL J. Cell Biol. 108:1177-1178(1989).
RN [3]
RP SEQUENCE OF 1695-2153 FROM N.A.
RA MEDLINE: 85284928.
RA Wasenius V.-M., Saraste M., Knowles J., Virtanen I., Lehto V.-P.;
RT "Sequencing of the chicken non-erythroid spectrin cDNA reveals an
RT internal repetitive structure homologous to the human erythrocyte
RT spectrin.";
RL EMBO J. 4:1425-1430(1985).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 965-1025.
RA Musacchio A., Noble M., Pauplit R., Wierenga R., Saraste M.;
RA "Crystal structure of a src-homology 3 (SH3) domain.";
RL Nature 359:851-855(1992).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 969-1025.
RA MEDLINE: 98363217.
RA Martinez J.C., Pisabarro M.T., Serrano L.;
RT "Obligatory steps in protein folding and the conformational diversity
RT of the transition state.";
RL Nat. Struct. Biol. 5:721-729(1998).
RN [6]
RP STRUCTURE BY NMR OF 1763-1872.
RA MEDLINE: 98022917.
RA Pascual J., Pfuhl M., Walther D., Saraste M., Nilges M.;
RT "Solution structure of the spectrin repeat: a left-handed
RT antiparallel triple-helical coiled-coil.";
RL J. Mol. Biol. 273:740-751(1997).
RN [7]
RP STRUCTURE BY NMR OF 2320-2403.
RA Trave G., Lacombe J.-P., Pfuhl M., Saraste M., Pastore A.;
RT "Molecular mechanism of the calcium-induced conformational change in
RT the spectrin EF-hands.";

```

```

RL EMBO J. 14:4922-4931(1995).
CC -1- FUNCTION: MORPHOLOGICALLY, SPECTRIN-LIKE PROTEINS APPEAR TO BE
CC RELATED TO SPECTRIN, SHOWING A FLEXIBLE ROD-LIKE STRUCTURE.
CC THEY CAN BIND ACTIN BUT SEEM TO DIFFER IN THEIR CALMODULIN-
CC BINDING ACTIVITY. IN NONERYTHROID TISSUES, SPECTRINS, IN
CC ASSOCIATION WITH SOME OTHER PROTEINS, MAY PLAY AN IMPORTANT
CC ROLE IN MEMBRANE ORGANIZATION.
CC -1- SUBUNIT: LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS
CC ARE CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO
CC TETRAMERS.
CC -1- DOMAIN: SPECTRIN-LIKE PROTEINS HAVE FIVE DOMAINS:
CC (1) N-TERMINAL DOMAIN (N),
CC (2) DOMAIN BETWEEN THE N-TERMINAL AND MIDDLE DOMAIN (NM),
CC (3) MIDDLE DOMAIN (M),
CC (4) DOMAIN BETWEEN THE MIDDLE AND C-TERMINAL DOMAIN (MC),
CC (5) C-TERMINAL DOMAIN (C).
CC NM AND MC DOMAINS ARE COMPOSED OF TYPICAL SPECTRIN 106 RESIDUES
CC REPEATS (1-8 FOR NM AND 12-19 FOR MC) AND ARE HOMOLOGOUS TO EACH
CC OTHER. N, M, AND C DOMAINS ARE COMPOSED OF SEQUENCES THAT DO NOT
CC FORM TYPICAL SPECTRIN REPEATS.
CC -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X14518; CAA32662.1; -.
CC EMBL: X14519; CAA32663.1; ALT_SEQ.
CC EMBL: X02593; CAB51571.1; ALT_SEQ.
CC PIR: A30122; STCHA.
CC PDB: 1SHG; 31-OCT-93.
CC PDB: 1AEY; 15-MAY-97.
CC PDB: 1AU3; 07-JUL-97.
CC PDB: 1TUC; 01-AUG-96.
CC PDB: 1TUD; 01-AUG-96.
CC PDB: 1BK2; 16-FEB-99.
CC PFAM: PF00018; SH3; 1.
CC PFAM: PF00036; efnand; 2.
CC PFAM: PF00435; spectrin; 23.
CC PRINTS: PR00452; SH3DOMAIN.
CC PROSITE: PS00018; EF_HAND; 2.
CC PROSITE: PS50002; SH3; 1.
CC KW Cytoskeleton; Membrane; Calmodulin-binding; Actin-binding;
CC Capring protein; Calcium-binding; Duplication; Repeat; SH3 domain;
CC 3D-structure.
CC KW
CC FT DOMAIN 1 14
CC FT REPEAT 15 119
CC FT REPEAT 120 225
CC FT REPEAT 226 331
CC FT REPEAT 332 437
CC FT REPEAT 438 543
CC FT REPEAT 544 648
CC FT REPEAT 649 754
CC FT REPEAT 755 860
CC FT REPEAT 861 966
CC FT DOMAIN 967 1061
CC FT DOMAIN 1062 1226
CC FT DOMAIN 1227 1309
CC FT REPEAT 1310 1415
CC FT REPEAT 1416 1521
CC FT REPEAT 1522 1633
CC FT REPEAT 1634 1739
CC FT REPEAT 1740 1845
CC FT REPEAT 1846 1951
CC FT REPEAT 1952 2058
CC FT REPEAT 2059 2171
CC -----
CC N-TERMINAL DOMAIN.
CC 1.
CC 2.
CC 3.
CC 4.
CC 5.
CC 6.
CC 7.
CC 8.
CC 9.
CC MIDDLE DOMAIN.
CC SH3.
CC MC DOMAIN.
CC 11.
CC 12.
CC 13.
CC 14.
CC 15.
CC 16.
CC 17.
CC 18.
CC 19.
CC 20.

```

FT REPEAT 2172 2256 21.
 FT DOMAIN 2257 2477 C-TERMINAL DOMAIN.
 FT CA_BIND 2341 2352 SITE 1 (POTENTIAL).
 FT CA_BIND 2384 2395 SITE 2 (POTENTIAL).
 FT STRAND 971 974
 FT STRAND 976 978
 FT TURAN 983 984
 FT STRAND 985 985
 FT STRAND 988 988
 FT TURAN 990 991
 FT STRAND 993 998
 FT STRAND 1004 1009
 FT TURAN 1010 1011
 FT STRAND 1012 1017
 FT HELIX 1018 1020
 FT STRAND 1021 1024
 SQ SEQUENCE 2477 AA; 285361 MW; ADAC876994E6AB39 CRC64;

Query Match
 Best Local Similarity 7.6%; Score 163.5; DB 1; Length 2477;
 Matches 100; Conservative 80; Mismatches 178; Indels 199; Gaps 18;

23 EVYLRLLDGRMKCVDELLRQRAAEERYGKELVQIARAGGOTEMNSLRTPDLS-- 80
 503 EALLKKHEDFEKSLAAGE--KITALDEFATKLIQ-----NNHYAMDVAATRDALIS 553
 81 -KOOTENVSAHIALALALREELRL--EEFEREKEORKKYEALMDRYOK 128
 554 RNALHERAMKRRQALDSFHLQOFFRDSDELKSWNKMKATDAVEAYKDPNSNOGKVK 613
 129 SK-----LSYKTTMESKAYDOK 147
 614 HQAFEAELSNQSRIDALEKAGOKLIDVNHASDEVAARMNEVISTLMKLEATELGKIK 673
 148 CRDADDAEQAEER-----VSANGHOKOVESONKAKOKESATEAEVYR 192
 674 LREANO--OQGFNRYVEDIELMVEEGHLSADYCKDLTSVQNLQK--KHALLADVAAH 730
 193 QN-----IEOLERATEWEOEHRTTCFAFLOEED 222
 731 QDPIDGITIAROFQDGHADNIRKKQALVARYEALDPMVARKOKLADSLRLQOLF 790
 223 RLTLIRNALV-----HCNLSQCVKDE 247
 791 R-DIEDETIIRKEPIAASTNRGKDLIGVQLKKHQALQALAEIAHEPRIKAVTQKNA 849
 248 LYEVRLLTEGCGVEGDIENG--IOSKSGREPPA--PVYONYTD-----R 290
 850 MVEEGHFAAE--DYKIKLMEINOKMDSLAKKASQORRODLEDLSQAOQYFADANEAQSMR 907
 291 EYVPLIGSPISQSCVIRFSGLLHSGSPKTPSPAPASTETLTPPERNELVYASIEVQ 350
 908 EKPIVIGSTD-----YKDEDSAEMLKKHKL-----MSDSLAVGSSISQ 947
 351 ATQGNLNSAODYR-----ALDYTAONSDELIDISAGDILAVILEGEGWMTV 398
 948 ALREQAOSCRQOVALPDDETGEKELVALDYQEKSPREVTMKGDILILNSTNKMWMKV 1007
 399 ERNORGVEVGSYLEKL 415
 1008 EYVNDROGFEVPAVYVKKL 1024

RESULT 10
 NCF2_BOVIN STANDARD; PRT; 527 AA.
 AC 077775;
 DT 15-FEB-2000 (Rel. 39, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last sequence update)
 DE NEUTROPHIL CYTOSOL FACTOR 2 (NCF-2) (NEUTROPHIL NADPH OXIDASE FACTOR 2) (P67-PHOX).

GN NCF2.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Burger P.L., Siemsen D., Quinn M.T.,
 RT "Cloning and expression of bovine NADPH oxidase proteins p47-phox and
 RT p67-phox.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: NCF2, NCF1, AND A MEMBRANE BOUND CYTOCHROME B558 ARE
 CC REQUIRED FOR ACTIVATION OF THE LATENT NADPH OXIDASE (NECESSARY
 CC FOR SUPEROXIDE PRODUCTION).
 CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -I- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AF079303; AAC82463.1; -
 DR HSSP: P08631; 1BU1.
 DR PFAM: PF00018; SH3; 2.
 DR PFAM: PF00515; TPR; 1.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR PRINTS: PR00499; P67PHOX.
 DR PROSITE: PS00002; SH3; 2.
 KW SH3 domain.
 FT DOMAIN 240 299 SH3.
 FT DOMAIN 458 517 SH3.
 SQ SEQUENCE 527 AA; 59663 MW; 430D33FE1DF68BD1 CRC64;

Query Match
 Best Local Similarity 7.5%; Score 162; DB 1; Length 527;
 Matches 69; Conservative 45; Mismatches 97; Indels 94; Gaps 12;

139 ESKRAYDQK--RDADDAEQAEFERSANGHOKOVESONKAKOKESATE----- 186
 55 EAEKAFKTSINRDNHNAVSYQRMGL---YYQMKYSAIKDKKALTLQRLQGLIDYKI 111
 187 -----AERYKQNIQELERATEWEOEHRTTCFAFLOEEDRLTLIRNALVHCNQ 237
 112 LGDQFKLFAECVLT--NIAFMYAKREEMKRAEHLALAVSMKSEPRHSKIDRA----- 162
 238 LSMQCVDDDELYEVRLLTEGCGVEGDIENGFIQSKSGREPPAPVYONYT--DREYTP 294
 163 --MESVWKQKLYE-----PVYIVGRILFRNEKQVNO 192
 295 LI-----GSPISQSCVIRFSGLLHSGSPKTPSPAPASTETLTP--TERNELVYASTE 348
 193 LVKKDYLGRKATVAASVDDPSFSGF-----APLQPAQAEPPRRKTP-----IFRALE 241
 349 VQATQGNLNSAODYRALDYTAONSDELIDISAGDILAVILEGEGWMTVERNGRGFVP 408
 242 GEA-----HVLGFGVPEPTEELQVMPGNIVFLKGNNDNMAIVENGGKGLVP 290
 409 GSYLE 413
 291 CNYLE 295

RESULT 11
 SPNC_HUMAN STANDARD; PRT; 2472 AA.
 AC 013813; Q16606;
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

01-NOV-1997 (Rel. 35, last annotation update)
 DE SPECTRIN ALPHA CHAIN, BRAIN (SPECTRIN, NON-ERYTHROID ALPHA CHAIN)
 GN SPTA2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 90170948.
 RA Moon R.T., McMahon A.P.;
 RT "Generation of diversity in nonerythroid spectrins. Multiple
 RT polypeptides are predicted by sequence analysis of cDNAs encompassing
 RT the coding region of human nonerythroid alpha-spectrin.";
 RT J. Biol. Chem. 265:4427-4433(1990).
 RN [2]
 RP SEQUENCE OF 676-1595 FROM N.A.
 RC TISSUE=LUNG;
 RX MEDLINE: 88083942.
 RA McMahon A.P., Moon R.T.;
 RT "Structure and evolution of a non-erythroid spectrin, human
 RT alpha-fodrin.";
 RT Biochem. Soc. Trans. 15:804-807(1987).
 RN [3]
 RP SEQUENCE OF 676-1595 FROM N.A.
 RX MEDLINE: 87277023.
 RA McMahon A.P., Giebelhaus D.H., Champion J.E., Bailes J.A., Lacey S.,
 RA Carritt B., Henchman S.K., Moon R.T.;
 RT "cDNA cloning, sequencing and chromosome mapping of a non-erythroid
 RT spectrin, human alpha-fodrin.";
 RT Differentiation 34:68-78(1987).
 RN [4]
 RP ERRATUM.
 RA McMahon A.P., Giebelhaus D.H., Champion J.E., Bailes J.A., Lacey S.,
 RA Carritt B., Henchman S.K., Moon R.T.;
 RL Differentiation 34:241-241(1987).
 CC -1- FUNCTION: FODRIN, WHICH SEEMS TO BE INVOLVED IN SECRETION,
 CC INTERACTS WITH CALMODULIN IN A CALCIUM-DEPENDENT MANNER AND IS
 CC CYTOSKELETON FOR THE CALCIUM-DEPENDENT MOVEMENT OF THE
 CC -1- SUBUNIT. LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS
 CC ARE CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO
 CC TETRAMERS.
 CC -1- DOMAIN: SPECTRIN-LIKE PROTEINS HAVE FIVE DOMAINS:
 CC (1) N-TERMINAL DOMAIN (N'),
 CC (2) DOMAIN BETWEEN THE N-TERMINAL AND MIDDLE DOMAIN (NM),
 CC (3) MIDDLE DOMAIN (M),
 CC (4) DOMAIN BETWEEN THE MIDDLE AND C-TERMINAL DOMAIN (MC),
 CC (5) C-TERMINAL DOMAIN (C).
 CC NM AND MC DOMAINS ARE COMPOSED OF TYPICAL SPECTRIN 106 RESIDUES
 CC REPEATS (1-8 FOR NM AND 12-19 FOR MC) AND ARE HOMOLOGOUS TO EACH
 CC OTHER. N, M, AND C DOMAINS ARE COMPOSED OF SEQUENCES THAT DO NOT
 CC FORM TYPICAL SPECTRIN REPEATS.
 CC -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: J05243; AAA51790.1; -
 CC EMBL: M24773; AAA52468.1; -
 CC EMBL: M18627; AAA51702.1; -
 CC HSSP: P07751; IAEY.
 CC MIM: 182810; -
 CC PFAM: PF00018; SH3; 1.
 CC PFAM: PF00036; efhand; 2.
 CC PFAM: PF00435; spectrin; 23.
 CC PRINTS: PR00452; SH3DOMAIN.

DR PROSITE; PS00018; EF_HAND; 1.
 DR PROSITE; PS50002; SH3; 1.
 KW Cytoskeleton; Membrane; Calmodulin-binding; Actin-binding;
 KW Capping protein; Calcium-binding; Duplication; Repeat; SH3 domain.
 FT DOMAIN
 FT REPEAT 1 14 N-TERMINAL DOMAIN.
 FT REPEAT 15 119
 FT REPEAT 120 225
 FT REPEAT 226 331
 FT REPEAT 332 437
 FT REPEAT 438 543
 FT REPEAT 544 648
 FT REPEAT 649 754
 FT REPEAT 755 860
 FT REPEAT 861 966
 FT REPEAT 967 1061
 FT REPEAT 1062 1251
 FT REPEAT 1062 1251 MC DOMAIN.
 FT REPEAT 1062 1167
 FT REPEAT 1204 1309
 FT REPEAT 1310 1415
 FT REPEAT 1416 1521
 FT REPEAT 1522 1628
 FT REPEAT 1629 1734
 FT REPEAT 1735 1840
 FT REPEAT 1841 1946
 FT REPEAT 1947 2053
 FT REPEAT 2054 2166
 FT REPEAT 2167 2251
 FT REPEAT 2252 2472 C-TERMINAL DOMAIN.
 FT CA_BIND 2336 2347 SITE 1 (POTENTIAL).
 FT CA_BIND 2379 2390 SITE 2 (POTENTIAL).
 FT CONFLICT 1595 1595 F -> R (IN REF. 2 AND 3).
 SQ SEQUENCE 2472 AA: 284279 MW: CDA1964462A7028D CRC64;

Query Match 7.5%; Score 161; DB 1; Length 2472;
 Best Local Similarity 18.1%; Pred. No. 0.043;
 Matches 81; Conservative 81; Mismatches 163; Indels 122; Gaps 16;

QY 44 RORQAERVKELVQIRKAGCO--TEMNSLRFSFSLKQOTENVGSAHQLALALAE 101
 DB 625 QSRIDALEKAGOKLIDVNHVAKDEVAARMNEVISLWKLLATE----- 668
 QY 102 LRSLEEFERQEKQKKEALMDRVQKSLLYKTKESKRAYDQCRDADAE-----Q 156
 DB 669 ---LKGITKLRANQOQQRN--NVEDLEWLYE--VGHSLASDDYGDNLNVQWQKH 720
 QY 157 AFEVVSANGHOKQYERKSONKAKCKESA--TEAERYRON---TEQLERARTWEQEHRTT 212
 DB 721 ALLEADVAAHQDRIDGVYIARQFODAGHFDENIKKQKQALVARYEALKEPVARKQKL 780
 QY 213 CEARQLOEFQDLTLIRNLWY-----HCNQ 237
 DB 781 ADSLRLOQLFR-DVEDEFTREKEPIAASNRGKDLGVNLKKHQAQAEIAGHPR 839
 QY 238 LSMQCVKDDLEVEFRVRLTEGCDVEGDINGF-----IOSKSTGHPRA--PVRYONYD 289
 DB 840 IKATQKGNMAYVEGHPAAE--DVKAKHELNQKWEALKAASQROLESISQAQGYFA 897
 QY 290 -----REVTPILGSPSIQPSGVIKRFSGILHSGPKTPPSAPASTETLPTPPERN 340
 DB 898 DANEAESMRREKEPIVGTST-----YKDEDSAEALLKKHEAL-----MS 937
 QY 341 ELVYASIEVQATQGNLWSSADYR-----ALYDTAQNSDLDLDSAGDI LAVI 388
 DB 938 DLSAYGSSIQALRQDAOSCRQOAVPTDDETKELVALDYQKSPREVTMKKGDIITLL 997
 QY 389 LEGEDGWTVERNGQRGFVPGSYLEKL 415
 DB 998 NSTKKDMWKVENVDRQGFVPAAYVKL 1024

RESULT 12

YB65_SCHPO STANDARD: PRT: 642 AA.

ID YB65_SCHPO

AC 009746:

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE HYPOTHETICAL 72.2 KDA PROTEIN C12C2.05C IN CHROMOSOME II.

GN SPB1C2.05C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;

OC Schizosaccharomycetaceae; Schizosaccharomyces.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-972;

RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;

RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.

CC -1- SIMILARITY: TO YEAST YHR114W.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: Z54140; CAA90818.1; -.

DR HSSP: P29354; 1GPD.

DR PFAM: PF00130; DAG_PE-bind; 1.

DR PFAM: PF00611; FCH; 1.

DR PFAM: PF00018; SH3; 2.

DR PRINTS: PR00008; DAGPEDOMAIN.

DR PROSITE: PS00479; DAG_PE_BINDING_DOMAIN; 1.

DR PROSITE: PS50002; SH3; 2.

DR Hypothetical protein; SH3 domain.

KW DOMAIN 521 575 SH3.

FT DOMAIN 584 642 SH3.

SEQUENCE 642 AA; 72216 MW; E6770ED15BD73D3C CRC64;

Query Match 7.5%; Score 160.5; DB 1; Length 642;
Best Local Similarity 16.7%; Pred. No. 0.0094;
Matches 107; Conservative 82; Mismatches 192; Indels 259; Gaps 20;

19 HTGEVLLQRLDGRKCKDVEELLROAQAERYKELVQIARKAG----- 65

13 HDKRYVDSWNNKAKMEDIQLYKKERSIEKRYAOKLASISNKGKSKSSALSVG 72

66 -----GQEMNSLRISFDSLKQOTENVGSAHIQL-----ALALKEELSLSEEFERQKE 114

73 DTPAMASGSLKASLI-TTWSKILDELTRSSKTHQKLSIDSDYSLDAIEKLKLESHTIE--A 128

115 QKKTYEALMDRVQSKSLYKKTMSKRAYQOKCHDADDAQAQAPERSVANGHQVQESQ 174

129 LRKYYDDLYKFFSSKEKTLNLSVRAKAVSYHACDDLESANQKNDKYEQKTRNLKLSIE 188

175 NKAQCKES-----ATEAERYRONIEQLERARTEWEOEHRTT-----CEAFQLO- 219

189 SDMLDKKKKYLRLMYVAHKQKFNELPLPLN-HMQVLMEYRNSNLTWCNSFSIEK 247

220 -----EPDRLTI- 226

248 SLHDTLSQRTVEIQSEIAKNEPVLDSAMFGRRHNSKNMALPADLHEPSPIMHDDDALVVD 307

227 -----LRNALMVAHCNOLSMQ---CYKDELYEYR-----LTLEEC- 259

308 GSKKNYLRNLVHSHKNDGKQKGLVSLDSQLEGLRVDDPNSANQSFESKASINLECKE 367

260 -----DVEGDIN----- 266

368 LMYKARIEDELEVRINKITSVANNNLEGGRFHDFKHVSFKLFTSCSYCEITWGLSKRCV 427

QY 267 ---GF-----IOSKSTGREP-----PAPVQXNYDYRE 291

DB 428 CKNGCFKCHARCCELLVYANCKNGEPEVADDVATSVATFDDPASASSNAEYSEYNT 487

QY 292 VTPILGSPSTQPSGVIKRPSGLHSGPKTPRSPAPASTELTLP---TPRENELYVAS-- 346

DB 488 TDDMDSSSYTQTSLSNKK-----TEETTPAPASKVGDVLYDFTGHEGYITASEG 539

QY 347 -----LEVQATQG-----NLNSAQDYRALVDYTAQN 373

DB 540 QEFTLLERDDSGSGVNRKIOCTDGLIPASYKLNDELNTSVTLDDGDSYKALAYTAQS 599

QY 374 SDELDISAGDILAVI-LEGEDGAWTVERNRGQFVPGSYL 412

DB 600 DMELSIOEGDILIQYTNRNAGNGWSEGLNGVTGQFPANYV 639

RESULT 13

YB65_SCHPO STANDARD: PRT: 777 AA.

ID YB65_SCHPO

AC 009697:

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE HYPOTHETICAL 88.2 KDA PROTEIN C2F7.18C IN CHROMOSOME I.

GN SPAC2F7.18C OR SPAC13A11.01C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;

OC Schizosaccharomycetaceae; Schizosaccharomyces.

RN [1]

RP SEQUENCE OF 1-147 FROM N.A.

RC STRAIN-972;

RA Hunt S., Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;

RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 114-777 FROM N.A.

RC STRAIN-972;

RA Gentles S., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;

RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: Z54096; CAA90802.1; -.

DR EMBL: Z50142; CAA90505.1; -.

DR HSSP: P01888; IBMG.

DR PFAM: PF00610; DEP; 1.

DR PFAM: PF00611; FCH; 1.

DR PFAM: PF00620; RhogAP; 1.

KW Hypothetical protein.

SEQUENCE 777 AA; 88207 MW; 707F64DCADEDAE6 CRC64;

Query Match 7.5%; Score 160.5; DB 1; Length 777;
Best Local Similarity 18.1%; Pred. No. 0.012;
Matches 109; Conservative 75; Mismatches 169; Indels 249; Gaps 20;

7 FRDAFWCRDFAHNGYVLLQRLDGRKCKDVEELLROAQAERYKELVQIARKA-- 64

DB 5 FSNQFWKSDVA--TGKKLFEDCLDNGVENEQYKNLKYKREANERGERGELQETTKCKLK 62

QY 65 GGQTEMN-----SLRTSPDSLKQOTENVGSAHIQALALR-----EELRSLSEEF 108

DB 63 GKRPENEDGATSKKAREGLRSEIANQCKOHIRAKDLETLILAFPSKMSIDHSQKQTS 122

QY 109 REKQKQKTYEALMDRVQSKSLYKKTMSKRAYQOKCHDADDAQ----- 156

Db 123 QOVLNTOIKSYE-----KKYTLAKT---KSAIYNKCNLDEYEESEKSESTTSEA 171
 QY 157 ----- 156
 Db 172 ITDLTVSSPQOQSLLENDODLIQLGFMEFPEELKEVLAQVLOEIPLODYRVPLIGTYP 231
 QY 157 -----A 157
 Db 232 NTCGNIIVSWLOENLEVPPLVAAEAFGODLIAQGLRHMVGSGSFVNSTNHFYQWKDKA 291
 QY 158 FERSANGHOKOYE-----KSONKAKOCKESATEAEVYRONI 195
 Db 292 FQFAGLSVDSLVENAKALPLVGYLSDYISHRKLVSSEYOSRKLREVLDAKNTYSESV 351
 QY 196 EOLERATEWEDEHRTTCFAFQLOEPRL-----TILNNALVHNCNLSMOCVK 244
 Db 352 VDLDKCTIVETIADHLQFLQKCESTRVLYKXDFMDLSTIINFL-----PSMKLLA 405
 QY 245 DD-ELYEVRLLTEGCGVEGDINGFIQSKSTGREPPAVPYQNYVDREVPPLIGSPSIQ 303
 Db 406 DOJIVYOEI-----IOPESDIRILESAATGPPLPREIYEDYVN-DIKDOI----- 451
 QY 304 SCGVYKFEGLLHSGPKTTP---SAPASSTETLTPPERNELVYASTEVQATQGNLN--- 357
 Db 452 -FGDVEF--LSHRDKRVRPIVSTILSYLDLPT-----LASDEVRONIMLVNSPL 501
 QY 358 SSAQDYBALDYTAQNSDEL-----DISAGDILAVILEGEDGWMVTERNGRGFVPSY 411
 Db 502 SSVHQLEALNHSSSVYKEVLSQYTPSVVIGVLKLYLELPD-----SIVPSA 550
 QY 412 LE 413
 Db 551 FE 552

RESULT 14
 ID R167_YEAST STANDARD: PRT: 482 AA.
 AC P39743;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 REDUCED VIABILITY UPON STRAVATION PROTEIN 167.
 GN RVS167 OR YDR388W OR D9509.8.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X2180;
 RX MEDLINE: 9330299.
 RA Bauer F., Urdaci M., Aigle M., Crouzet M.;
 RT "Alteration of a yeast SH3 protein leads to conditional viability
 with defects in cytoskeletal and budding patterns";
 RL Mol. Cell. Biol. 13:5070-5084(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E.,
 RA Berno A., Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
 RA Huntke-Smith S., Hymen R., Komp C., Lashkari D., Lew H., Lin D.,
 RA Moseedale D., Nakahara K., Nemath A., Oefner P., Oh C., Petel F.X.,
 RA Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,
 RA Winant A., Yelton M., Botstein D., Davis R.W.;
 RL Submitted (Jul-1995) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP ACTIN-BINDING.
 RX MEDLINE: 95236199.
 RA Amberg D.C., Basart E., Botstein D.;
 RT "Defining protein interactions with yeast actin in vivo";
 RL Nat. Struct. Biol. 2:28-35(1995).
 CC -!- FUNCTION: COMPONENT OF A CYTOSKELETAL STRUCTURE THAT IS REQUIRED
 FOR THE FORMATION OF ENDOCYTIC VESICLES AT THE PLASMA MEMBRANE
 CC LEVEL. COULD BE IMPLICATED IN CYTOSKELETAL REORGANIZATION IN

CC RESPONSE TO ENVIRONMENTAL STRESSES AND COULD ACT IN THE BUDDING
 CC SITE SELECTION MECHANISM. BINDS TO ACTIN.
 CC -!- SIMILARITY: TO YEAST RVS161.
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M92092; AAA35051.1; -;
 DR EMBL: U32274; AAB64830.1; -;
 DR PIR: S40887; S40887.
 DR HSP: P04002; 1A7F.
 DR SGD: L0001789; RVS167.
 DR PFAM: PF00018; SH3.1.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR PROSITE: PS50002; SH3; 1.
 KM Cytoskeleton: SH3 domain; Transmembrane; Actin-binding.
 FT DOMAIN 292 427 ALA/GLY/PRO-RICH.
 FT TRANSMEM 344 367 POTENTIAL.
 FT DOMAIN 421 482 SH3.
 SQ SEQUENCE 482 AA; 52774 MW; 3F0AB53EBC95A5B CRC64;

Query Match 7.4%; Score 159.5; DB 1; Length 482;
 Best Local Similarity 19.2%; Pred. No. 0.0077;
 Matches 98; Conservative 75; Mismatches 194; Indels 143; Gaps 20;

QY 3 AOLFRDAFWCRDFTAHNGYEVILQRLDGRKMGCDVEELLRORAOAEERKELVQJAR 62
 Db 13 APQSTRQKFKNGEOTDEYVE-----DAERFPELOETKSLSESKRY----- 56
 QY 63 KAGQTEHNSLRTSPDSLKQOTENVSAHIQALALRELSLEEFRRQKORRYEAI 122
 Db 57 -----STAVNGMLTHQIGFAKSMEEIFK--PISGKSDPNMTIPDNPQGIASQYRAI 109
 QY 123 MDRVQK-----KLSLYKKTKMESKKAYDKCRADDAEQAFERVS 162
 Db 110 VAELOETLKPDLALVEEKIVTPCQELKITYIRKMAKR--NHRKLDRLHNTY----- 163
 QY 163 ANGHQOYKESQONKAKOCKESATEAEVYRONIEQLERARTEWEGEH----- 209
 Db 164 -NKHKKKE-----PTAKDEERLYKAO-AQVEVAQOETDIYNDLKTQLPILFSL 211
 QY 210 -----RTTCEAFQLOEPRLTILRNALN---VHCNQLSMOCVKD-----DEL- 248
 Db 212 EAEFVKPLFEVFYFQNLIFTLVRLQDMKIPYFDLMSDIVESIYAKNGVEQTDALT 271
 QY 249 -----YEEVRL-----TLEGCDVEGDINGFIQSKSTGREP-----PADVPYQN 286
 Db 272 ITHFKLGYSKAKLEMTRRKRYGATAGSPVSGASGV--GYGAGDYPATATSPPTGY-G 328
 QY 287 Y-----YDREVTPLIGSPSIQPSGVTKRFGSLHGS-PKTPPAPAASTETTL----- 334
 Db 329.YGAAPSTAAOPAAOYGTAAAVGTAAAGTAAGAAVPGYTPQYTAQAQSPPLTGLGFQ 388
 QY 335 -----PPERNELVYASTEVQATQGNLNSAODYRALYDYTAQNSDELIDISAGDIL 385
 Db 389 QSPQOQGPFPAYSNPLTSPVAGTPAAVAAPVGETYALTYDQAQAAGLSFPAGAVI 448
 QY 386 AVILEGED--GMWYVERNGRGFVPGSTYLE 413
 Db 449 EIVQRTPDVNMETGGRYNGQOGVFPGNVYQ 478

RESULT 15
 SRC8_CHICK STANDARD: PRT: 563 AA.
 ID SRC8_CHICK
 AC 001406;


```

67 nfhrglumetlasSerLeuArgThrSerPheaspSerLeuLysGlnGlnT 84
|||||
882 GACAGAGATGAATTCCTGAGGACCTCTTGACTCCTCGAAGACACAA 931
|||||
84 hrgluasnValGlySerAlaHisIleGlnLeuAlaLeuArgGlu 100
|||||
932 CAGAGAAATGGGCAGTGCACACATCCAGCTGCTGGCCCTGCTGAG 981
|||||
101 GtuleuArgSerLeuGlnGluPheArgGluArgGlnLysGlnGlnArg 117
|||||
982 GAGCTGGGAGCTGGAGAGATTCCGAGAGACACAAAGAGAGGAGAA 1031
|||||
117 slsYrGluAlaIleMetAspArgValGlnLysSerLysLeuSerLeuT 134
|||||
1032 GAAGTATGAGGCCATCATGAGCCGTGTCACAAAGACAGATTGCTGCT 1081
|||||
134 yrlsLysThrMetGluSerLysAlaTyrAspGlnLysCysArgAsp 150
|||||
1082 ACAAGAAAGACATGGAGTCCAAAGAGCATATGACCAGAAAGTGCAGG 1131
|||||
151 AlaaspAspAlaGlnGlnAlaPheGluArgValSerAlaAsnGlyHisG 167
|||||
1132 GCAGATGATGCTGAGCAGGCTTCGAGCGTGTGAGTGCACATGGCCACA 1181
|||||
167 nlysGlnValGluLysSerGlnAsnLysAlaLysGlnCysGluSerA 184
|||||
1182 GAAGCAAGTAGAAAGAGCCACAGAACCAAGCAGTGCACAGAGATCAG 1231
|||||
184 lATHrgluAlaGluArgValTyrArgGlnAsnIleGlnLeuGluArg 200
|||||
1232 CCACAGAGGACAGAAAGATGTACAGCAAAATATCCAACTGGGAGAA 1281
|||||
201 AlaArgThrGluTropGluGlnLysArgGlyThrCysGluAlaPheG 217
|||||
1282 GCGAGAGCCGAGTGGAGACAGAGACCCGAGTACTGAGAGCCCTTCCA 1331
|||||
217 nleuGlnGluPheAspArgLeuThrIleLeuArgAsnAlaLeuTPrValH 234
|||||
1332 GTTGCAGGATTGTGACCGGCTCACCATCTCCGCAATGCCCTGTGGCTC 1381
|||||
234 lAcCysAsnGlnLeuSerMetGlnCysValLysAspAspGluLeuTyrGlu 250
|||||
1382 ACTGTAAACCAAGCTCTCCATGAGTGTGTCAAAGATCATAGCTTAATG 1431
|||||
251 GluValArgLeuThrLeuGlnGlyCysAspValGluGlyAspIleAsnG 267
|||||
1432 GAAGTGGGCTGACCTTGAGGGCTGTGATGTGGAAGTGCATCAATG 1481
|||||
267 yPheIleGlnSerLysSerThrGlyArgGluProProAlaProValProT 284
|||||
1482 CTTCATCCAGTCCAAAGACACTGGCAGAGAGCCGCCAGCTCGGTCCT 1531
|||||
284 yrgLAsnTyrTyrAspArgGluValThrProLeuIleGlySerProSer 300
|||||
1532 ATCCAAACTACTATGACAGGAGGTGACCCCACTAATTTGGACCCCTTAC 1581
|||||
301 lIleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisG 317
|||||
1582 ATCCAGCCCTCTGCGGTGTGATAAAGAGTCTCTGGGCTCTACATG 1631
|||||
317 ySerProLysThrThrProSerAlaProAlaIleSerThrGluThrLeuT 334
|||||
1632 AAGTCCCAAGACACACACTTGTGCTCTGCTCTCCACAGAGACTCTGA 1681
|||||
334 hrProThrProGluArgAsnGluLeuValTyrAlaSerIleGluValGln 350
|||||
1682 CTCACACCCCTGAGCGGAATGAGTTGTCTACGATCCATCCAAAGTGCAG 1731
|||||
351 AlaThrGlnGlyAsnLeuAsnSerSerAlaGlnAspTyrArgAlaLeuVal 367
|||||
1732 GCGACCCAGGAAACCTTACTCATCAGCCGACGACTACCGGAGACTCTA 1781
|||||

```

```

367 rAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspI 384
|||||
1782 CGACTTACACTGCACAGAAATTTGATGAGCTGACATTTCCGGGAGACA 1831
|||||
384 lIleuAlaValIleLeuGlnGlyGluAspGlyTyrTrpThrValGluArg 400
|||||
1832 TCCTGGCGGTCAATCCTGGAAAGGAGAGATGCTGTGTGACTGTGAGCGG 1881
|||||
401 AsnGlyGlnArgGlyPheValProGlySerTyrLeuGlnLysLeu 415
|||||
1882 AACGGACAAACGCTGCTTGTCTGCTGGTGTGATCTTGACAAAGCTC 1926
|||||

```

seq_name: N_Geneseq_36: r39809

seq_documentation_block:

ID r39809 standard; DNA; 4091 BP.

AC r39809; 23-FEB-1998 (first entry)

DE Mouse H74 gene.

KW Src-homology region 3 domain; human; mouse; SH3 domain; cell growth; cellular signalling element; cellular structural element; malignancy;

KW protein identification; functional domain; protein screening; cellular signal transduction process; ss.

OS Mus musculus.

FS Location/Qualifiers

FH Key

FT CDS

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

alignment_scores:

Quality: 262.50

Ratio: 1.014

Percent Similarity: 54.989

Percent Identity: 21.019

alignment_block:

Sequence 4091 BP; 945 A; 1232 C; 1116 G; 798 T;

US-09-068-377-1 x T39809 ..

Align seg 1/1 to: T39809 from: 1 to: 4091

```

3 AlaGlnLeuGlnPheArgAspAlaPheThrPysArgAspPheThrAlaH1 19
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
221 GCCACAGAGGATCACAGATAGCTTCTGG.....GGA 252
19 sThrGly...TyrGluValLeuLeuGlnArgLeuLeuAspGlyArgLysM 35
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
253 GGTGGGGAACACACAGCGAGCGTGAACGCCATGACGATGGGCCACCCCC 302
35 eTcysLysAspValGluGluLeuLeuArgGlnArgAlaGlnAlaGluL1 51
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
303 TGTCAACAGACCTCATGAGCTGGCTGCAGAGCGGCCCAAGATGACAGAAG 352
52 ArgGlyGlyLysGluLeuValGlnIleAlaArgLys..... 63
353 GCATACGCGGACGAGCTCACCGACTGGGCGCAAGCGCTGGCCGACCTCAT 402
64 ....AlaGlyGlnThrGluMetAsnSerLeuArgThrSerPheAsp 79
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
403 CGAATAAGGTCCTCAG.....TATGGCAGCCTGAGCGGGCGCTGGGCG 446
79 eLeuLysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAla 95
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
447 CCATGATGACAGACAGATTAAGGTCAACGAGCTGCACACGAGAGGTGAAG 496
96 LeuAlaLeuArgGluGluLeuArgSerLeuGluGluPheArgGluArgG1 112
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
497 AACGCTGCTGTAATGAG.....GACCTGGAGAAATCAAGAACTGGCA 540
112 nLysGlu.....GlnArgLysL1 118
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
541 GAAGGATGGCCATACACAAGCATATCGGTGCTTCAAGAGACAGAAAG 590
118 ySTyrgLualIaIeMetAspArgValGlnLysSerLysLeuSerLeuTyr 134
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
591 AGGCGGAGATGGCTCCGAAAGGCCGGAAGCCCTGGCTGCTAAAGAAATG 640
135 LysLysThrMetGluSerLysLysAlaTyrAspGlnLysCysArgAspAl 151
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
641 AAGGAGCTAGAGCGCGCCAAAGAGCCTATCATTTGGCTTAAAGAGGA 690
151 AspAspAla.....GluGlnAlaPheGluArgValSerAlaAsnGln 166
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
691 AAGGCTGGCCATGACCCGGGAGATGAACAGTAAGACAGACAGATCGGTCA 740
166 IsGlnLysGlnValGluLysSerGlnAsnLysAlaLysGlnCysLysGlu 182
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
741 CCCCTGAACAGACAGAAACTTGTGCAAAAGTGCACAAATGCGACAG 790
183 SerAlaThrGluAlaGluArgValTyrArgGlnAsnIleGluGlnLeuG1 199
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
791 GATGTGCAAAAGACTCAGAGAACTATGAGAAAGTCTCGGAAAGATGTGG 840
199 uATGAlaArgThrGluTyrGluGlnGluHisArgThrCysGluAlaP 216
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
841 CAAGACCAACACACAGTACATGGAGGCGATGACAGAGTGTGGAGAGT 890
216 heGlnLeuGlnGluPheAspArgLeuThrIleLeuArg..... 228
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
891 GCCAGCAGTTTGAAGAGAGCGCTGCTCTCTGCAAGGAAGTCTCTCTG 940
229 .....AsnAlaLeuThrPValHisCys 235
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
941 GATATCAAAAGGCATCTCAACCTAGCGAGAACAGACAGCTACATGCAT.. 988
235 sAsnGlnLeuSerMetGlnCysValLysAspAspGluLeuTyrGlnGlu 252
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
989 .....GTCTACCGAGAAC 1001
252 aLArgLeuThrLeuGluGlyCysAspValGluGlyAspIleAsnGlyPhe 268

```

```

1002 TCGACGAGGCGCATCCGGGGCGGATGCCAGAGGAGCCTCAGGTGTTTC 1051
269 ILeGlnSerLysSerThrGlyArgGluProProAlaProValProTyrG1 285
1052 CCGACGACCATGTCGCCCGGAGT.....CCATGAACTGGCGCAGATT 1095
285 nAsnTyrTrpAspArgGluValThrProLeuIleGlySerProSerIleG 302
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1096 CGAGGAGTGGAAACCACTCCCGACACACACATGCGCAAGAGAAAC 1145
302 InProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySer 318
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1146 ACCCT.....AAGAGCGAGAGGG..... 1165
319 ProLysThrThrProSerAlaProAlaAlaSerThrGluThrLeuThr 335
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1166 .....GCCACCTGAGCAATGCCACTGGGGCTGTAGAAATCCACATCCCA 1209
335 cThrProGluArgAsnGluLeu.....ValTyrA 345
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1210 GGCTGGGAGACCTGGCAGTGTATGACAGCTATGACGAGCGCAACATATG 1259
345 lAserIleGluValGlnAlaThrGlnGlyAsn..... 355
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1260 CCACGAGGTGTACAGACATGAGAGCGGAACCCCTTGGGGCAATGAG 1309
356 LeuAsnSerSerAlaGlnAspTyr..... 363
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1310 GCCAATGTGGCGCCAAACCCCTTCGAGATGATGCCAAGGAGTTGCTGT 1359
364 .ArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIle 380
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1360 ACGGGCAGCTGTATGACTAGACGTGAGGAGCAGATGAGCTCAGCTTCA 1409
380 eTAlaGlyAspIleLeuAlaValIleLeuGluGlyGluAspGlyTyrTP 396
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1410 ACGCGGAGATGAGTCAACCAAGCTCGGAGAGAGAACAGACAGGGTGG 1459
397 ThrValGluArg.....AsnGlyGlnArgGlyPheValProGlySerTyr 411
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1460 TCCCGCGGGCGGCTGGACACGCGAGCAGCTGGGCTCTATCTGCCAATTA 1509
411 rLeuGluLysLeu 415
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1510 CTTTGAACGTATA 1522

seq_name: N_Geneseq_36:T39810
seq_documentation_block:
ID T39810 standard; DNA; 1133 BP.
AC T39810;
DT 23-FEB-1998 (first entry)
DE Human H74 gene.
KW Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
cellular signalling element; cellular structural element; malignancy;
KW protein identification; functional domain; protein screening;
KW cellular signal transduction process; ss.
OS Homo sapiens.
FH Key 1. 1133
FT CDS
FT /transl_except= (pos: 1..2, aa: Arg)
FT /note= "this codon has an apparent 1 nucleotide deletion,
which alters the reading frame"
WO9631625-A1.
PD 10-OCT-1996.
PF 04-APR-1996; U04454.
PR 03-APR-1996; US-630915.
PA 07-APR-1995; US-417872.
PA (CYTO-) CYTOGEN CORP.
PA (UYN-) UNIV NORTH CAROLINA.
PA Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;

```

alignment_scores:

alignment_block:

US-09-068-377-1 x T39810

Align seg 1/1 to: T39810 from: 1 to: 1133

[illegible]

cellular signalling element; cellular structural element; malignancy; SH3 domain; human; mouse; cell growth; SH3 domain; human; mouse; cell growth; malignancy; cellular signalling element; cellular structural element


```

2524 CCAGCAGGAGAGCTACCCAGACAGAGCTGACCCCTCATGATTTGTT 2573
362 . . . AspTyrArgAlaLeuTyrAspTyrThrAlaIlnsSerAspGluLeu 377
2574 GTAGCTACCAAGCGTTATATAGTTATGTGCACAGAACATGATGAGTTG 2623
378 AspIleSerAlaGlyAspIleLeuAlaValIleLeuGluGlyGluAsp 394
2624 GAACTCCGAGATGAGATATGTTGATGATGAGAAAAATGACGATGG 2673
394 yTrpTrp.....ThrValGluArgAsnGlyIlnArgGlyPheValProG 409
2674 ATGCTTTGTTGGCACTTCGACGAAGGACGAGCGATTTGTACTTTCCAG 2723
409 ySerTyrLeuGluIlnGlyLeu 415
2724 GCACATATGTAAACCTTTA 2743

seq_name: N_Geneseq_36.T39789

seq_documentation block:
ID T39789 standard; DNA; 1302 BP.
AC T39789.
DE 18-FEB-1998 (first entry)
DE Mouse SH3P7 gene.
KW Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
KW cellular signalling element; cellular structural element; malignancy;
KW protein identification; functional domain; protein screening;
KW cellular signal transduction process; ss.
OS Mus musculus.
FH Key
FT CDS
FT 1. 1302 Location/Qualifiers
FT /*tag= a
FT /product= SH3P7
PN WO9631625-A1.
PD 10-OCT-1996.
PE 04-APR-1996; U04454.
PR 03-APR-1996; US-630915.
PR 07-APR-1995; US-417872.
PA (CYTO-) CITROGEN CORP.
PA (UNIC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;
DR WPJ; 96-465045/46.
P-PSDB; W05389.
PT Identifying polypeptide(s) having specific functional domain (esp.
PT SH3 domain) comprises detecting selective binding to recognition
PT unit, regardless of sequence homology
PS Claim 76: Fig 30; 174pp; English.
CC T39786-T39803 represent novel human and mouse genes encoding Src-homology
CC region 3 (SH3) domain containing proteins that can be used in the method
CC of the invention. SH3 domain containing proteins play a role in
CC signalling and structural elements of cells. The method of the invention
CC is for identifying polypeptides containing functional domains of interest
CC (especially SH3 domains). The method comprises contacting a multivalent
CC polypeptides having a selective binding affinity for the RU complex. The
CC method is based on functional similarities and does not rely on sequence
CC similarities. Prior methods only gave limited success for identifying
CC proteins which contain an SH3 domain due to the minimal sequence homology
CC among known SH3 proteins. It has been found that small peptide RUs in
CC multivalent form have reduced specificity for a given functional domain
CC compared to monomer RUs. Multivalent RU complexes are particularly suited
CC to screening for polypeptides containing functional domains that are
CC similar to, but not identical in sequence to, the original target
CC functional domain. The new method enables proteins having a common
CC useful for a better understanding of cell growth, malignancy, signal
CC transduction processes, etc. New candidate drugs can be identified, and
CC their specificities (e.g. pharmacological activities) can be assessed
CC using the method of the invention.
SQ Sequence 1302 BP; 328 A; 359 C; 412 G; 203 T;

```

```

alignment_scores:
Quality: 179.50 Length: 397
Ratio: 0.930 Gaps: 14
Percent Similarity: 48.615 Percent Identity: 20.151

alignment block:
US-09-068-377-1 x T39789 ..

Align seg 1/1 to: T39789 from: 1 to: 1302

47 AlaGlnAlaGluGlnArgTyrGlyGlySerGluLeuVal.....GlnIleAl 61
|||||
349 GCCCGGGCCGAGAGAGGTGAGAGCCCTGATGATCATGAGGAAGTTGC 398
61 aArgLysAlaGlyGlyGlnThrGlnMetAsnSerLeuArgThrSerPheA 78
|||||
399 CAAGGCTCTGGGGCCCACTACAGCTTCCATAGAGAAAGCACTCTCTCC 448
78 sPserLeuLysGlnGlnThrGlnAsnValGlySerAlaHisIleGlnLeu 94
|||||
449 AGGATGTAGGGCCGAGGCCCA...GTGGCTCTGTGTAC..... 486
95 AlaLeuAlaLeuArgGluGluLeuArgSerLeuGluIlnPheArgGluAr 111
|||||
487 .....CAGAACCAATGCCATCTGATGATCAAG..... 516
111 gGlnLysGluIlnArgLysTyrGluAlaIleMetAspArgValGlnL 128
|||||
517 .....AGAGTGGGCA 526
128 ySerLeuLysSerLeuTyrLysLysThrMetCysSerLysLysAlaTyr 144
|||||
527 AGGATTACTTCTGGGCCMAAGCTGAGAGAGAGAGAAAGAACCGCGCTCG 576
145 AspGlnLysCysArgAspAlaAspAlaSerPheGlnAlaPheGluArgVa 161
|||||
577 GAGAGAGAGCGCGCT.....GCCGAAGAGAGACGGCAGCGGCTT 614
161 lSerAlaAsnGlyHisGlnLysGlnValGlnLysSerGlnAsnLysAlaL 178
|||||
615 GAGAGAGAGCGAGAGAGCGGAGAGCTGAGAGAGCTGCCGACGTGAC 664
178 yGlnCysLysGlu.....SerAlaThrGluAlaGluArgValTyr 191
|||||
665 AGGCTACACAGAGACGACAGATGAGTGGACCCGACGAGACAGCT 714
192 ArgGlnAsnIleGluGlnLeuGluAlaArgAlaArgThrGluTrpGluGln 208
|||||
715 GAGCCAGACAGAGAGAGCGTTTCAAGGACGACAGAGAGTGGAG..... 759
208 uHisArgThrThrCysGluAlaPheGlnLeuGlnGluPheAspArgLeuT 225
759 .....PheIleG 270
225 hIleLeuArgAsnAlaLeuTyrValHisCysAsnGlnLeuSerMetGln 241
760 .....TCTGTGGGACGAG 774
242 CysValLysAspAspGluLeuTyrGluGluValAlaArgLeuThrLeuGlu 258
|||||
775 GCCCCACACCCAGAGAGATTTTCAAGCAGAGAAAGGCAATGTCCAC 824
258 yCysAspValGluGlyAspIleAsnGly.....PheIleG 270
825 CACCTGTTCACCAAGCTCGAGCGGGCAAGCTGAGAGACCCCTTCCTGC 874
270 lSerLysSerThr.....GlyArgGluProProAla 280
|||||
875 AGAAGCAATCACTCAACAGAAACCTCTACGGCCGAGAGCCACAGCT 924
281 ProValProTyrGlnAsnTyrTyrAspArgGluValThrProLeuIleGl 297
925 CCGTCTCC..... 933

```



```

297 yserProSerIleGlnProSerCysGlyValIleYsArgPheSerGlyL 314
      ::::::::::::::::::::
934 .....CGGCTGACAGCAGGTGCTGTGAG..... 957
314 euLeuHisGlySerProLysThrThrProSerAlaProAlaAlaSerThr 330
      ::::::::::::::::::::
958 .....GAGCCACGCGCTACGACTGTCTGTGCGCCAGACAGAAAGAA 1002
331 GluThrLeuThrProThrProGluArgAsnGluLeuValTyrAlaSerIle 347
      ||| |||
1003 CCTACATATGAAAGTACCCCAAGACGACACCCTC...TATGAGAACCC 1049
347 egluValGlnAlaThrGlnGlyAsnLeuAsnSerAlaGlnAspTyr 363
      ::::::::::::::::::::
1050 ACCACTGTGACAGCAGCAAGGGGCTGGCTCCGACACATGACAACTACA 1099
364 .....ArgAlaLeuTyr 367
      |||
1100 TGCAGAGCCAGGCTTCACTGACAGCAAGGGCTGTGCGCCGCGCTGTAT 1149
368 AspTyrThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspIle 384
      ||||| ||| ::::::::::::::::::::
1150 GACTACACAGGCGAGCTGATGACACCGAGATCTCTTGACCTGAGAACCT 1199
384 eLeuAlaValIleLeuGlnGlyLeuAspGlyTyrTrpThr...ValGluA 400
      ::::::::::::::::::::
1200 AATCACAGGACATCGAGGTGATGAGCAAGCGTGTGCGAGGCTATGGGC 1249
400 rGAsnGlyGlnArgGlyPheValProGlySerTyrLeuGlu 413
      ::::::::::::::::::::
1250 CTGACGGCCACTTGTGGCATGTCTTCCGCCACTACTGTGAG 1290
seq_name: N_Geneseq_36:T39791

```

```

seq_documentation_block:
ID T39791 standard; DNA; 1531 BP.
AC T39791;
AT 18-FEB-1998 (first entry)
DE Mouse SH3P9 gene.
KW Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
cellular signalling element; cellular structural element; malignancy;
protein identification; functional domain; protein screening;
KW cellular signal transduction process; ss.
OS Mus musculus.
FH Key location/Qualifiers
FT CDS 1..1302
FT FT /*tag=a
FT FT trans_except= (pos: 1475..1477, aa: Val)
FT FT /product= SH3P9
PN W0631625-A1.
PD 10-OCT-1996.
PE 04-APR-1996; U04454.
PR 03-APR-1996; US-630915.
PR 07-APR-1995; US-417872.
PA (CYTO-) CYTOGEN CORP.
PA (UYN-) UNIT NORTH CAROLINA.
PI Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;
DR WBI: 96-465045/46.
DR P-PSDB: W05391.
PT Identifying polypeptide(s) having specific functional domain (esp.
PT SH3 domain) - comprises detecting selective binding to recognition
PT unit, regardless of sequence homology
PS Claim 76; Fig 34; 174pp; English.
CS T39786-T39803 represent novel human and mouse genes encoding Src-homology
region 3 (SH3) domain containing proteins that can be used in the method
of the invention. SH3 domain containing proteins play a role in
CC signalling and structural elements of cells. The method of the invention
CC is for identifying polypeptides containing functional domains of interest
CC (especially SH3 domains). The method comprises contacting a multivalent
CC recognition unit (RU) complex with a number of peptides and identifying
CC polypeptides having a selective binding affinity for the RU complex. The
CC method is based on functional similarities and does not rely on sequence

```

CC similarities. Prior methods only gave limited success for identifying CC proteins which contain an SH3 domain due to the minimal sequence homology CC among known SH3 proteins. It has been found that small peptide RUs in CC multivalent form have reduced specificity for a given functional domain CC compared to monomer RUs. Multivalent RU complexes are particularly suited CC to screening for polypeptides containing functional domains that are CC similar to, but not identical in sequence to, the original target CC functional domain. The new method enables proteins having a common CC function to be identified. Identification of novel SH3 proteins will be CC useful for a better understanding of cell growth, malignancy, signal CC transduction processes, etc. New candidate drugs can be identified, and CC their specificities (e.g. pharmacological activities) can be assessed CC using the method of the invention.

alignment_scores: Quality: 173.50 Length: 457
Ratio: 0.778 Gaps: 19
Percent Similarity: 48.796 Percent Identity: 21.225

alignment_block:
US-09-068-377-1 x T39791 ..

Align seq 1/1 to: T39791 from: 1 to: 1531

```

33 ArgLysMetCysLysAspValGluGluLeuArgGlnArgAlaGlnAl 49
      ::::::::::::::::::::
236 AAGAGCTGACCCGACGCCAGAGAGAGAGTCTTGCAAGAACTGGGAAAGC 285
49 agluGluArgTyrGlyGlyLeuValGlnIleAlaArgLysAlaGly 66
      ::::::::::::::::::::
286 GAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 335
66 LysGln...ThrGlu.....MetAsnSerLeuArgThrSerPhe 77
      ||| |||
336 AGCAGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 385
78 AspSerLeuLysGlnGlnThrGluAsnValGlySerAlaHisIleGln 94
      ||| |||
386 GCTTCTGTTAAAGCGATCGACGAA..... 409
94 uAlaLeuAlaLeuArgGluGluLeuArgSerLeuGluGluPheArgGlu 111
      ::::::::::::::::::::
410 .....GCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 452
111 rglGlnLysGlnGlnArgLysLysTyrGluAlaIle..... 122
      ::::::::::::::::::::
453 CCGAGTGGCCTGGCAGGAGATGAAGCAAGATTGCGAGAACATGAC 502
123 .....MetAspArgValGlnLysSerLysLeuSerLeuTyrLysLys 136
      ||| |||
503 CTACTCTGGATGGACTACACACAGAG.....CTGGTGGACCA 540
136 strMetGluSerLysLysAlaTyr.....AspGlnLysC 148
      ::::::::::::::::::::
541 GCGTCTGCTGACCATGACACCTACCTAGGACAGTCCGATATCAAGT 590
148 ysrGAspAlaAspAspAlaGluGlnAlaPheGluArgValSerAlaAsn 164
      ::::::::::::::::::::
591 CCGGCAATGGCAAGCGGGGGGAGAGTGTGACTGTGACAGTGGCCGG 640
165 GlnHisGlnLysGlnValGlu.....LysSerGlnAsnLysAl 177
      ||| |||
641 CACCACATATAGTCTTTCAAACCGCCAAAAGAGATGAAGCCAAAT 690
177 alvGlnCysLysLysSerAlaThrGluAlaGluArgValTyrArgGln 193
      ::::::::::::::::::::
691 TGCACAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 740
194 ...AsnIleGlnGlnLeuGluArgAlaArgThrGluTyrGlnGlnHis 209
      ||| |||
741 TGAACGTGATTCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 790

```

```

210 ArgThrThrCysGluAlaPheGlnLeu..... 218
      :::::|||||
791 GGTTCATGTCAACAGCTTCCAGACATCGCGGCTCGAGAAACTT 840
219 ....GlnGluPheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValH 234
      :::::|||||
841 CCATTAAGAGATAGTAGTAACTCATCAGAACCTCAATGATGCTCGTCA 890
234 As.....CysAsnGlnLeuSerMetGlnCysVallys 244
      |||
891 GCCTAGAGAGACAGACGAGGAGCAACACCTTCACAGTCAAGCCCAACCC 940
245 AspAspGluLeuTrpGlu..... 250
      :::::|||||
941 AGTGACAATGCCCTGAGAAAGGAAAGAGCCCGTCACTCTCTCCAGA 990
251 .....GluValArgLeuThrLeuGlnGlyCysA 260
991 TGGCTCCCTGCTGCTACCTGAGATCAGATGATCAAGCAGAG..... 1033
260 spValGlnGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyArg 276
1034 .....CCAGAGCGGCGCAGTGGGCGC 1054
277 GluProPheAlaProValProTyrGlnAsnTyrTyrAspArgGluValTh 293
1055 TCACCCGGGCTACCATCCCAAGTCCCATTCACAGCCAGACAGAGCGCTTC 1104
293 rProLeuIleGlySerProSerIleGlnProSerCysGlyValIleLysA 310
1105 CGAGGTGGTGGTGGAGCCAG...GAGCCA..... 1132
310 rPheSerGlyLeuLeuHisGlySerProLysThrThrProSerAlaPro 326
1133 .....GGGAGACAGACGCCAGCAAGCAACCTCCAGCTCTCTTCG 1174
327 AlaIleSerThrGluThrLeuThrProThrProGluArgAsnGluLeuVal 343
1175 GCGTGGTGGTGGAGACCTTCCGCACT.....CT 1206
343 lTyraIleSerIleGluValGlnAlaThrGlnGlyAsnLeuAsn..... 357
1207 GAATGGGGGGTGGAGGCGAGCGTGGAGCTGGAGCTGGAGCTGGAGCC 1256
358 ..SerSerIleGlnAspTyrArgAlaLeuTyrAspTyrThrAlaGlnAsn 373
1257 CGGAGTTTCATGTTCAGGTTCAAGCCACAGATATTACACGCGCACTGAC 1306
374 SerAspGluLeuAspIleSerAlaGlyAspIleLeuAlaValIle..... 388
1307 ACTGATGAGCTGCACACTCAAGCTGCGCATGTGCTTGATTCCTTT 1356
389 .....LeuGlnGlyGluAspGlyTrp..... 395
1357 CCAGAACCCAGAGAGAGAGAGATGAGGATGAGTGCATGAGTGTGAAGAGA 1406
396 .....TrpThrValGluArgAsnGlyGln.....ArgGlyPheValPro 408
1407 GCGAGTGGATTCAGACACAGAAAGAACTGAGAAATGCCGCGCTTCCCG 1456
409 GlySerIleGlnGluLysLeu 415
1457 GAGAAATTTTACAGAGCGGCTA 1477

seq_name: N_Geneseq_36:T39805

seq_documentation_block:
ID T39805 standard; DNA; 1867 BP.
AC T39805:
DT 20-FEB-1998 (first entry)
DE Mouse cortactin gene.
KW Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;

```

```

KW cellular signalling element; cellular structural element; malignancy;
KM protein identification; functional domain; protein screening;
KW cellular signal transduction process; ss.
OS Mus musculus.
FH key
FT Location/Qualifiers
FT CDS
FT 131..1771
FT /*tag=a
   /transl_except= (pos: 1283..1285, aa: Glu)

PD WO9631625-A1.
PD 10-OCT-1996.
PF 04-APR-1996; U04454.
PR 03-APR-1996; US-630915.
PR 07-APR-1995; US-417872.
PA (CYTO-) CYTOGEN CORP.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;
DR WPJ: 96-465045/46.
DR P-PSDB; W05406.
PT Identifying polypeptide(s) having specific functional domain (esp.
PT SH3 domain) - comprises detecting selective binding to recognition
PT unit, regardless of sequence homology
PS Claim 102; Fig 26; 174pp; English.
CC T39804-T39810 represent human and mouse genes encoding Src-homology
CC region 3 (SH3) domain containing proteins that can be used in the method
CC of the invention. SH3 domain containing proteins play a role in
CC signalling and structural elements of cells. The method of the invention
CC is for identifying polypeptides containing functional domains of interest
CC (especially SH3 domains). The method comprises contacting a multivalent
CC recognition unit (RU) complex with a number of peptides and identifying
CC polypeptides having a selective binding affinity for the RU complex. The
CC method is based on functional similarities and does not rely on sequence
CC similarities. Prior methods only gave limited success for identifying
CC proteins which contain an SH3 domain due to the minimal sequence homology
CC among known SH3 proteins. It has been found that small peptide RUs in
CC multivalent form have reduced specificity for a given functional domain
CC compared to monomer RUs. Multivalent RU complexes are particularly suited
CC to screening for polypeptides containing functional domains that are
CC similar to, but not identical in sequence to, the original target
CC functional domain. The new method enables proteins having a common
CC function to be identified. Identification of novel SH3 proteins will be
CC useful for a better understanding of cell growth, malignancy, signal
CC transduction processes, etc. New candidate drugs can be identified, and
CC their specificities (e.g. pharmacological activities) can be assessed
CC using the method of the invention.
Sequence 1867 BP; 505 A; 437 C; 570 G; 355 T;

alignment_scores:
Quality: 164.00 Length: 396
Ratio: 0.953 Gaps: 15
Percent Similarity: 43.434 Percent Identity: 22.980

alignment_block:
US-09-068-377-1 x T39805 ..

Align seg 1/1 to: T39805 from: 1 to: 1867

40 GlnGluLeuLeuArgGlnArgAlaGlnAlaGlnGluGluArgGlyLysG1 56
|||||
890 GAGAGCTGCAGCTCATGAAATCCCAAAAAGCATATGAGTGGT..... 934
56 uLeuValGlnIleAlaArgLysAlaGlyGlnGlnThrGluMetAsnSerI 73
935 .....TTGCGAGGCAAAATTTGGTGTCCAGTCCGAG..... 964
73 euArgThrSerPheAspSerLeuLysGlnGlnThrGluAsnValGlySer 89
965 .....AGGAGAGACTCTCCGCTCGGTGTT 991
90 AlaHisIleGlnLeuAlaLeuAlaLeuArgGluGluLeuArgSerLeuG1 106
992 GATTAC.....AAGGAGAGATTGGCCCAAGCAGCA 1020

```

```

106 uGluPheArgGluArgGlnLysGlnArgLysLysTyrGluAlaIleW 123
107      :      :      :      :      :      :      :      :      :
108      :      :      :      :      :      :      :      :      :
109      :      :      :      :      :      :      :      :      :
110      :      :      :      :      :      :      :      :      :
111      :      :      :      :      :      :      :      :      :
112      :      :      :      :      :      :      :      :      :
113      :      :      :      :      :      :      :      :      :
114      :      :      :      :      :      :      :      :      :
115      :      :      :      :      :      :      :      :      :
116      :      :      :      :      :      :      :      :      :
117      :      :      :      :      :      :      :      :      :
118      :      :      :      :      :      :      :      :      :
119      :      :      :      :      :      :      :      :      :
120      :      :      :      :      :      :      :      :      :
121      :      :      :      :      :      :      :      :      :
122      :      :      :      :      :      :      :      :      :
123      :      :      :      :      :      :      :      :      :
124      :      :      :      :      :      :      :      :      :
125      :      :      :      :      :      :      :      :      :
126      :      :      :      :      :      :      :      :      :
127      :      :      :      :      :      :      :      :      :
128      :      :      :      :      :      :      :      :      :
129      :      :      :      :      :      :      :      :      :
130      :      :      :      :      :      :      :      :      :
131      :      :      :      :      :      :      :      :      :
132      :      :      :      :      :      :      :      :      :
133      :      :      :      :      :      :      :      :      :
134      :      :      :      :      :      :      :      :      :
135      :      :      :      :      :      :      :      :      :
136      :      :      :      :      :      :      :      :      :
137      :      :      :      :      :      :      :      :      :
138      :      :      :      :      :      :      :      :      :
139      :      :      :      :      :      :      :      :      :
140      :      :      :      :      :      :      :      :      :
141      :      :      :      :      :      :      :      :      :
142      :      :      :      :      :      :      :      :      :
143      :      :      :      :      :      :      :      :      :
144      :      :      :      :      :      :      :      :      :
145      :      :      :      :      :      :      :      :      :
146      :      :      :      :      :      :      :      :      :
147      :      :      :      :      :      :      :      :      :
148      :      :      :      :      :      :      :      :      :
149      :      :      :      :      :      :      :      :      :
150      :      :      :      :      :      :      :      :      :
151      :      :      :      :      :      :      :      :      :
152      :      :      :      :      :      :      :      :      :
153      :      :      :      :      :      :      :      :      :
154      :      :      :      :      :      :      :      :      :
155      :      :      :      :      :      :      :      :      :
156      :      :      :      :      :      :      :      :      :
157      :      :      :      :      :      :      :      :      :
158      :      :      :      :      :      :      :      :      :
159      :      :      :      :      :      :      :      :      :
160      :      :      :      :      :      :      :      :      :
161      :      :      :      :      :      :      :      :      :
162      :      :      :      :      :      :      :      :      :
163      :      :      :      :      :      :      :      :      :
164      :      :      :      :      :      :      :      :      :
165      :      :      :      :      :      :      :      :      :
166      :      :      :      :      :      :      :      :      :
167      :      :      :      :      :      :      :      :      :
168      :      :      :      :      :      :      :      :      :
169      :      :      :      :      :      :      :      :      :
170      :      :      :      :      :      :      :      :      :
171      :      :      :      :      :      :      :      :      :
172      :      :      :      :      :      :      :      :      :
173      :      :      :      :      :      :      :      :      :
174      :      :      :      :      :      :      :      :      :
175      :      :      :      :      :      :      :      :      :
176      :      :      :      :      :      :      :      :      :
177      :      :      :      :      :      :      :      :      :
178      :      :      :      :      :      :      :      :      :
179      :      :      :      :      :      :      :      :      :
180      :      :      :      :      :      :      :      :      :
181      :      :      :      :      :      :      :      :      :
182      :      :      :      :      :      :      :      :      :
183      :      :      :      :      :      :      :      :      :
184      :      :      :      :      :      :      :      :      :
185      :      :      :      :      :      :      :      :      :
186      :      :      :      :      :      :      :      :      :
187      :      :      :      :      :      :      :      :      :
188      :      :      :      :      :      :      :      :      :
189      :      :      :      :      :      :      :      :      :
190      :      :      :      :      :      :      :      :      :
191      :      :      :      :      :      :      :      :      :
192      :      :      :      :      :      :      :      :      :
193      :      :      :      :      :      :      :      :      :
194      :      :      :      :      :      :      :      :      :
195      :      :      :      :      :      :      :      :      :
196      :      :      :      :      :      :      :      :      :
197      :      :      :      :      :      :      :      :      :
198      :      :      :      :      :      :      :      :      :
199      :      :      :      :      :      :      :      :      :
200      :      :      :      :      :      :      :      :      :
201      :      :      :      :      :      :      :      :      :
202      :      :      :      :      :      :      :      :      :
203      :      :      :      :      :      :      :      :      :
204      :      :      :      :      :      :      :      :      :
205      :      :      :      :      :      :      :      :      :
206      :      :      :      :      :      :      :      :      :
207      :      :      :      :      :      :      :      :      :
208      :      :      :      :      :      :      :      :      :
209      :      :      :      :      :      :      :      :      :
210      :      :      :      :      :      :      :      :      :
211      :      :      :      :      :      :      :      :      :
212      :      :      :      :      :      :      :      :      :
213      :      :      :      :      :      :      :      :      :
214      :      :      :      :      :      :      :      :      :
215      :      :      :      :      :      :      :      :      :
216      :      :      :      :      :      :      :      :      :
217      :      :      :      :      :      :      :      :      :
218      :      :      :      :      :      :      :      :      :
219      :      :      :      :      :      :      :      :      :
220      :      :      :      :      :      :      :      :      :
221      :      :      :      :      :      :      :      :      :
222      :      :      :      :      :      :      :      :      :
223      :      :      :      :      :      :      :      :      :
224      :      :      :      :      :      :      :      :      :
225      :      :      :      :      :      :      :      :      :
226      :      :      :      :      :      :      :      :      :
227      :      :      :      :      :      :      :      :      :
228      :      :      :      :      :      :      :      :      :
229      :      :      :      :      :      :      :      :      :
230      :      :      :      :      :      :      :      :      :
231      :      :      :      :      :      :      :      :      :
232      :      :      :      :      :      :      :      :      :
233      :      :      :      :      :      :      :      :      :
234      :      :      :      :      :      :      :      :      :
235      :      :      :      :      :      :      :      :      :
236      :      :      :      :      :      :      :      :      :
237      :      :      :      :      :      :      :      :      :
238      :      :      :      :      :      :      :      :      :
239      :      :      :      :      :      :      :      :      :
240      :      :      :      :      :      :      :      :      :
241      :      :      :      :      :      :      :      :      :
242      :      :      :      :      :      :      :      :      :
243      :      :      :      :      :      :      :      :      :
244      :      :      :      :      :      :      :      :      :
245      :      :      :      :      :      :      :      :      :
246      :      :      :      :      :      :      :      :      :
247      :      :      :      :      :      :      :      :      :
248      :      :      :      :      :      :      :      :      :
249      :      :      :      :      :      :      :      :      :
250      :      :      :      :      :      :      :      :      :
251      :      :      :      :      :      :      :      :      :
252      :      :      :      :      :      :      :      :      :
253      :      :      :      :      :      :      :      :      :
254      :      :      :      :      :      :      :      :      :
255      :      :      :      :      :      :      :      :      :
256      :      :      :      :      :      :      :      :      :
257      :      :      :      :      :      :      :      :      :
258      :      :      :      :      :      :      :      :      :
259      :      :      :      :      :      :      :      :      :
260      :      :      :      :      :      :      :      :      :
261      :      :      :      :      :      :      :      :      :
262      :      :      :      :      :      :      :      :      :
263      :      :      :      :      :      :      :      :      :
264      :      :      :      :      :      :      :      :      :
265      :      :      :      :      :      :      :      :      :
266      :      :      :      :      :      :      :      :      :
267      :      :      :      :      :      :      :      :      :
268      :      :      :      :      :      :      :      :      :
269      :      :      :      :      :      :      :      :      :
270      :      :      :      :      :      :      :      :      :
271      :      :      :      :      :      :      :      :      :
272      :      :      :      :      :      :      :      :      :
273      :      :      :      :      :      :      :      :      :
274      :      :      :      :      :      :      :      :      :
275      :      :      :      :      :      :      :      :      :
276      :      :      :      :      :      :      :      :      :
277      :      :      :      :      :      :      :      :      :
278      :      :      :      :      :      :      :      :      :
279      :      :      :      :      :      :      :      :      :
280      :      :      :      :      :      :      :      :      :
281      :      :      :      :      :      :      :      :      :
282      :      :      :      :      :      :      :      :      :
283      :      :      :      :      :      :      :      :      :
284      :      :      :      :      :      :      :      :      :
285      :      :      :      :      :      :      :      :      :
286      :      :      :      :      :      :      :      :      :
287      :      :      :      :      :      :      :      :      :
288      :      :      :      :      :      :      :      :      :
289      :      :      :      :      :      :      :      :      :
290      :      :      :      :      :      :      :      :      :
291      :      :      :      :      :      :      :      :      :
292      :      :      :      :      :      :      :      :      :
293      :      :      :      :      :      :      :      :      :
294      :      :      :      :      :      :      :      :      :
295      :      :      :      :      :      :      :      :      :
296      :      :      :      :      :      :      :      :      :
297      :      :      :      :      :      :      :      :      :
298      :      :      :      :      :      :      :      :      :
299      :      :      :      :      :      :      :      :      :
300      :      :      :      :      :      :      :      :      :
301      :      :      :      :      :      :      :      :      :
302      :      :      :      :      :      :      :      :      :
303      :      :      :      :      :      :      :      :      :
304      :      :      :      :      :      :      :      :      :
305      :      :      :      :      :      :      :      :      :
306      :      :      :      :      :      :      :      :      :
307      :      :      :      :      :      :      :      :      :
308      :      :      :      :      :      :      :      :      :
309      :      :      :      :      :      :      :      :      :
310      :      :      :      :      :      :      :      :      :
311      :      :      :      :      :      :      :      :      :
312      :      :      :      :      :      :      :      :      :
313      :      :      :      :      :      :      :      :      :
314      :      :      :      :      :      :      :      :      :
315      :      :      :      :      :      :      :      :      :
316      :      :      :      :      :      :      :      :      :
317      :      :      :      :      :      :      :      :      :
318      :      :      :      :      :      :      :      :      :
319      :      :      :      :      :      :      :      :      :
320      :      :      :      :      :      :      :      :      :
321      :      :      :      :      :      :      :      :      :
322      :      :      :      :      :      :      :      :      :
323      :      :      :      :      :      :      :      :      :
324      :      :      :      :      :      :      :      :      :
325      :      :      :      :      :      :      :      :      :
326      :      :      :      :      :      :      :      :      :
327      :      :      :      :      :      :      :      :      :
328      :      :      :      :      :      :      :      :      :
329      :      :      :      :      :      :      :      :      :
330      :      :      :      :      :      :      :      :      :
331      :      :      :      :      :      :      :      :      :
332      :      :      :      :      :      :      :      :      :
333      :      :      :      :      :      :      :      :      :
334      :      :      :      :      :      :      :      :      :
335      :      :      :      :      :      :      :      :      :
336      :      :      :      :      :      :      :      :      :
337      :      :      :      :      :      :      :      :      :
338      :      :      :      :      :      :      :      :      :
339      :      :      :      :      :      :      :      :      :
340      :      :      :      :      :      :      :      :      :
341      :      :      :      :      :      :      :      :      :
342      :      :      :      :      :      :      :      :      :
343      :      :      :      :      :      :      :      :      :
344      :      :      :      :      :      :      :      :      :
345      :      :      :      :      :      :      :      :      :
346      :      :      :      :      :      :      :      :      :
347      :      :      :      :      :      :      :      :      :
348      :      :      :      :      :      :      :      :      :
349      :      :      :      :      :      :      :      :      :
350      :      :      :      :      :      :      :      :      :
351      :      :      :      :      :      :      :      :      :
352      :      :      :      :      :      :      :      :      :
353      :      :      :      :      :      :      :      :      :
354      :      :      :      :      :      :      :      :      :
355      :      :      :      :      :      :      :      :      :
356      :      :      :      :      :      :      :      :      :
357      :      :      :      :      :      :      :      :      :
358      :      :      :      :      :      :      :      :      :
359      :      :      :      :      :      :      :      :      :
360      :      :      :      :      :      :      :      :      :
361      :      :      :      :      :      :      :      :      :
362      :      :      :      :      :      :      :      :      :
363      :      :      :      :      :      :      :      :      :
364      :      :      :      :      :      :      :      :      :
365      :      :      :      :      :      :      :      :      :
366      :      :      :      :      :      :      :      :      :
367      :      :      :      :      :      :      :      :      :
368      :      :      :      :      :      :      :      :      :
369      :      :      :      :      :      :      :      :      :
370      :      :      :      :      :      :      :      :      :
371      :      :      :      :      :      :      :      :      :
372      :      :      :      :      :      :      :      :      :
373      :      :      :      :      :      :      :      :      :
374      :      :      :      :      :      :      :      :      :
375      :      :      :      :      :      :      :      :      :
376      :      :      :      :      :      :      :      :      :
377      :      :      :      :      :      :      :      :      :
378      :      :      :      :      :      :      :      :      :
379      :      :      :      :      :      :      :      :      :
380      :      :      :      :      :      :      :      :      :
381      :      :      :      :      :      :      :      :      :
382      :      :      :      :      :      :      :      :      :
383      :      :      :      :      :      :      :      :      :
384      :      :      :      :      :      :      :      :      :
385      :      :      :      :      :      :      :      :      :
386      :      :      :      :      :      :      :      :      :
387      :      :      :      :      :      :      :      :      :
388      :      :      :      :      :      :      :      :      :
389      :      :      :      :      :      :      :      :      :
390      :      :      :      :      :      :      :      :      :
391      :      :      :      :      :      :      :      :      :
392      :      :      :      :      :      :      :      :      :
393      :      :      :      :      :      :      :      :      :
394      :      :      :      :      :      :      :      :      :
395      :      :      :      :      :      :      :      :      :
396      :      :      :      :      :      :      :      :      :
397      :      :      :      :      :      :      :      :      :
398      :      :      :      :      :      :      :      :      :
399      :      :      :      :      :      :      :      :      :
400      :      :      :      :      :      :      :      :      :
401      :      :      :      :      :      :      :      :      :

```

```

1672 CATCACCAACATAGAAATGATGATGACGATGCTGCGTGGGTGCA 1721
401 sngGlnArgGlyPheValProGlySerTyrLeuIleu 413
1722 AGGGCAGATGAGGGCTCTTCCAGCCACTATGTGTGAG 1759

seq_name: N_Geneseq_36:T39802

seq_documentation_block:
ID T39802 standard; DNA; 549 BP.
AC T39802;
DT 19-FEB-1998 (first entry)
DE Human clone 53 gene.
KW Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
KW cellular signalling element; cellular structural element; malignancy;
KW protein identification; functional domain; protein screening;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..549
FT /tag= a
FT /transl_except= (pos: 343..345, aa: Gln)
FT /transl_except= (pos: 517..519, aa: Arg)
PE WO9631625-A1.
PD 10-OCT-1996.
PF 04-APR-1996: U04454.
PR 03-APR-1996: US-630915.
PR 07-APR-1995: US-417872.
PA (CYTO-) CYTOGEN CORP.
PI (UYN-) UNIV NORTH CAROLINA.
PI Fowkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;
DR WP: 96-465045/46.
DR P-PSDB: W05402.
PT Identifying polypeptide(s) having specific functional domain (esp.
PT SH3 domain) - comprises detecting selective binding to recognition
PT unit, regardless of sequence homology
PS Claim 76; Fig 64; 174pp; English.
CC T39786-T39803 represent novel human and mouse genes encoding Src-homology
CC region 3 (SH3) domain containing proteins that can be used in the method
CC of the invention. SH3 domain containing proteins play a role in
CC signalling and structural elements of cells. The method of the invention
CC is for identifying polypeptides containing functional domains of interest
CC (especially SH3 domains). The method comprises contacting a multivalent
CC recognition unit (RU) complex with a number of peptides and identifying
CC polypeptides having a selective binding affinity for the RU complex. The
CC method is based on functional similarities and does not rely on sequence
CC similarities. Prior methods only gave limited success for identifying
CC proteins which contain an SH3 domain due to the minimal sequence homology
CC among known SH3 proteins. It has been found that small peptide RUS in
CC multivalent form have reduced specificity for a given functional domain
CC compared to monomer RUS. Multivalent RU complexes are particularly suited
CC to screening for polypeptides containing functional domains that are
CC similar to, but not identical in sequence to, the original target
CC functional domain. The new method enables proteins having a common
CC function to be identified. Identification of novel SH3 proteins will be
CC useful for a better understanding of cell growth, malignancy, signal
CC transduction processes, etc. New candidate drugs can be identified, and
CC their specificities (e.g. pharmacological activities) can be assessed
CC using the method of the invention.
SQ Sequence 549 BP; 159 A; 127 C; 134 G; 129 T;

```

```

alignment_scores:
  Quality: 154.50      Length: 207
  Ratio: 1.644        Gaps: 9
  Percent Similarity: 45.411  Percent Identity: 26.087

```

alignment_block:

US-09-068-377-1 x T39802 ..

Align seg 1/1 to: T39802 from: 1 to: 549

```

222 AspaRgLeuThrIleLeuAlaGlnAlaLeuTrpValHisCysAsnGlnLeu 238
      ::::::::::::::::::::
118 GAGAGATCATCACTGCTCCGACAGGAA..... 144
238 uSerMetGlnCysValLysAspAspGluLeuTrpGluGluValArgLeuT 255
      |||:::|:::|:::|
145 .....GATGAGAACTGCTAC..... 159
255 hrLeuGluGlnCysAspValGluGlnLysAspIleAsnGlnPheIleGlnSer 271
      |||||:::|:::|
160 .....GAAGGAGGATCCCGGG..... 177
272 LysSerThrGlyArgGluProProAlaProValProTyrGlnAsnTyrTy 288
      |||:::|:::|
178 .....ACATCCCGACACAGCATCTTCCCATCCATCCATCCATGATGTGAT 221
288 rAspArgGluValThrProLeuIleGlySerPro..... 299
      |||:::|:::|
222 CAAGGGA.....CCACTGGTGAAACCCCTGTGATTACATGAGCC 262
      |||:::|:::|
300 .....SerIleGlnProSerCysGlyValIleLysArgPheSerGly 313
      |||:::|:::|
263 TGCCTTTCCTCCCTCCCAAGT..... 285
314 LeuLeuHisGlySerProLysThrThrProSerAlaProAlaLaserTh 330
      |||:::|:::|
286 .....CCGAGTCCACACTCCAAAGCCACAGCAACCTCAAGCCACGCA 326
330 rGluThrLeuThrProThrProGluArgAsnGlnLeuValTyrAlaSerI 347
      |||:::|:::|
327 GCGAAGAGTCACCC..... 342
347 LeGluValGlnAlaThrGlnGlnLysLeuAsnSerSerAlaGlnAsp... 362
      |||:::|:::|
343 .....GACAGAGTCACAAACCTCCACAGATTAA 369
363 .....TyrArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGluLe 377
      |||:::|:::|
370 TTTAGCTATCAAGCATATATAGCTATATACCAACAAATGATGATGAGTT 419
377 uAspIleSerAlaGlyAspIleLeuAlaValIleLeuGlnGluAspG 394
      |||:::|:::|
420 GGAATCTCCCGATGGAGATATCTGATGATCATGAAAAATGTGACGATG 469
394 LTTTPTP.....ThrValGluArgAsnGlnArgGlyPheValPro 408
      |||:::|:::|
470 GATGGTTTGTGTGACTTCACAGAGACAAAGCAGTTTGTACTTTTCCA 519
408 GlySerTyrLeuGlnLysLeu 415
      |||:::|:::|
520 GGCACACTATGTAACCTTTG 540
seq_name: N_Geneseq_36:T39798
seq_documentation_block:
ID T39798 standard; DNM; 1687 BP.
AC T39798;
DT 19-FEB-1998 (first entry)
DE Human clone 56 gene.
KW SRC-homology region 3 domain; human; mouse; SH3 domain; cell growth;
cellular signalling element; cellular structural element; malignancy;
protein identification; functional domain; protein screening;
KW cellular signal transduction process; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 3..1070
FT /tag= a
FT WO96J1625-A1.
PD 10-OCT-1996.
PD 04-APR-1996; U04454.
PR 03-APR-1996; US-630915.
PR 07-APR-1995; US-417872.
PA (CYTO-) CYTOGEN CORP.

```

```

PA (UTNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;
DR WPI: 96-465045/46.
DR P-PSDB: W05398.
PT Identifying polypeptide(s) having specific functional domain (esp.
SH3 domain) comprises detecting selective binding to recognition
unit, regardless of sequence homology
PS Claim 76; Fig 56; 174pp; English.
CC T39786-T39803 represent novel human and mouse genes encoding Src-homology
region 3 (SH3) domain containing proteins that can be used in the method
of the invention. SH3 domain containing proteins play a role in
signalling and structural elements of cells. The method of the invention
is for identifying polypeptides containing functional domains of interest
(especially SH3 domains). The method comprises contacting a multivalent
recognition unit (RU) complex with a number of peptides and identifying
polypeptides having a selective binding affinity for the RU complex. The
method is based on functional similarities and does not rely on sequence
similarities. Prior methods only gave limited success for identifying
proteins which contain an SH3 domain due to the minimal sequence homology
among known SH3 proteins. It has been found that small peptide RUs in
multivalent form have reduced specificity for a given functional domain
compared to monomer RUs. Multivalent RU complexes are particularly suited
to screening for polypeptides containing functional domains that are
similar to, but not identical in sequence to, the original target
functional domain. The new method enables proteins having a common
function to be identified. Identification of novel SH3 proteins will be
useful for a better understanding of cell growth, malignancy, signal
transduction processes, etc. New candidate drugs can be identified, and
their specificities (e.g. pharmacological activities) can be assessed
using the method of the invention.
SQ Sequence 1687 BP; 352 A; 486 C; 512 G; 337 T;

alignment_scores:
Quality: 153.00 Length: 292
Ratio: 1.070 Gaps: 9
Percent Similarity: 48.973 Percent Identity: 20.548

alignment_block:
US-09-068-377-1 x T39798
Align seg 1/1 to: T39798 from: 1 to: 1687

147 LysCysArgAspAlaAspAlaGlnGlnAlaPheGluArgValSerAl 163
      |||:::|:::|
135 AAAGCAGAGAAAGCAGAGCAGAAACCTGCGTGAAGAAACCGTGCGGC 184
163 AsnGlnHisGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnC 180
      |||:::|:::|
185 CGAGAGAGCGACGCGGACGCTGGACAG...GAGCGCGGAGCGGTGAGC 231
180 yAlysGlnSerAlaThrGlnAlaGluArgValTyrArgGlnIleGlu 196
      |||:::|:::|
232 TGCGTGAGGCTCAGCGCGGAGCAGCGCTATCAGAGACAGGGTGCGAG 281
197 GlnLeuGluArgAlaArgThrGlnTrpGlnGlnGlnHisArgThrThyCy 213
      |||:::|:::|
282 GCCACGCCCCAGAGCAGAGAGC...TGGGAGCAGCAGCAAGAGTGGTTTC 328
213 sGluAlaPheGlnLeuGlnGlnLysPheAspArgLeuThrIleLeuArgAsnA 230
      |||:::|:::|
339 AAGGAACCGAAATGAGCAGAGG..... 350
230 laLeuTrpValHisCysAsnGlnLeuSerMetGlnCysValLysAspAsp 246
      |||:::|:::|
351 .....TTGCGCGTGACCCCGAGG 368
247 GluLeuTyrGlnGluValArgLeuThrLeuGlnGlnCysAspValGlnI 263
      |||:::|:::|
369 GAGATTTCACAGCAGACAGAGGCGCATGTCCACACCTCATCTCCAG 418
263 yAspIleAsnGln.....PheIleGlnSerLysSerThr. 274
      |||:::|:::|

```

```

419 TCCTCAAGCTGGCAAGCTGAGAGACCCCTCTCTGACAGAGACGCTACCC 468
275 .....GlyArgLupProAlaProValProTyrGln 285
469 AACCAAGACCACTTGGCAGAGACCACTGCTGCATCTCA..... 512
286 AsnTyrTyrAspArgLupValTyrProLeuIleGlySerProSerIleG1 302
513 .....AGCCCAAGGACAGATCTCCCTGCTGAGCA 541
302 nProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerP 319
542 GCCG.....GGCG 549
319 rGlyThrThrProSerAlaProAlaAlaSerThrGluThrLeuThrPro 335
550 CCAGCACTCTCCATGTCGTGTGACAGCAGAGAGAGAGAGCTGTATGAG 599
336 ThrProGluArgAsnGluLeuValTyrAlaSerIleGluValGlnAlaTh 352
600 GAACCTCCAGACGAGAGACCTCTACGACGAGCCCACTGTGACAGCA 649
352 rGlnGly.....AsnL 356
650 GCAAGGTGCTGCTCTGACGACATTGACCAACATTGAGGCGCAGGCG 699
356 euAsnSerSerAlaGlnAspTyrArgAlaLeuTyrAspTyrThrAlaGln 372
700 TCAGTGTGGCAAGGCTCTGTCGCCCTGCTGCTGACACTACACAGGACCC 749
373 AsnSerAspLupLeuAspLysSerIleGlyAspIleLeuAlaValIleLe 389
750 GACGACACAGAGATCTCTTGTACCCGAGAACCTCATCAGCGCATGCA 799
389 uGluGlyLupAspLysTyrThrThr...ValGluArgAsnGluGlnArg 405
800 GGTGATGACAGAGAGCTGTGTGGCTGCTGATGAGCCGGATGTCATTG 849
405 LysPheValProGlySerTyrLeuGlu 413
850 GCATGTTCCCTGCCAATTACGTGAG 875

seq_name: N_Geneseq_36:T43573

seq_documentation_block:
ID T43573 standard; cDNA; 1925 BP.
AC T43573:
DT 07-FEB-1997 (first entry)
DE Human box-dependent myc-interacting protein (BIN1) cDNA.
KW Box-dependent myc-interacting protein; oncogene; oncoprotein;
tumor suppressor gene; cancer; hyperplasia; breast cancer;
KW liver cancer; probe; diagnosis; gene therapy; ds.
OS Homo sapiens.
RH key Location/Qualifiers
cDS 60..1415
FT misc-feature
FT 813..854
FT /tag= a
FT /note= "bases 813-854 code for BIN1 nuclear
localisation domain and are useful for
probe design"
FT misc-feature
FT 867..908
FT /tag= c
FT /note= "bases 867-908 code for BIN1 myc-binding
domain and are useful for probe design"
FT misc-feature
FT 891..1412
FT /tag= d
FT /note= "bases 891-1412 code for BIN1 SH3 domain
and are useful for probe design"
W09634627-A1.
PN 02-MAY-1996; U06231.
PP 05-MAY-1995; U5-435454.

```

```

PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Prendergast GC, Sakamuro D;
DR WPI: 96-505902/50.
DR P-PSDB: W06602
PT Mammalian box-dependent myc-interacting protein Bin1 coding
sequences - used in the diagnosis and treatment of c-myc associated
PT cancers and hyperplastic disease and Bin1 deficiency
PS Claim 1; Page 44-47; 79pp; English.
CC A cDNA clone (T43573) codes for the human box-dependent myc-
interacting protein Bin1 (W06602). Binding of Bin1 to c-Myc
CC appears to regulate c-Myc, resulting in tumour suppression. The
CC cDNA clone was isolated from a human skeletal muscle lambda-ZAPI
CC using murine Bin1 cDNA (see also T43574) as probe. A partial Bin1
CC genomic sequence (T43574) has also been obt. Bin1 nucleic acids
CC can be used to produce Bin1 proteins in host cells, or incorporated
CC into vectors for the gene therapy of Bin1 deficiency diseases and
CC cancers or hyperplastic diseases involving the c-myc oncogene.
CC Gene fragments are useful as probes for detecting these diseases.
SQ Sequence 1925 BP; 467 A; 534 C; 581 G; 343 T;

```

```

alignment_scores:
Quality: 150 50 Length: 461
Ratio: 0.660 Gaps: 20
Percent Similarity: 49.458 Percent Identity: 21.692

```

```

alignment_block:
US-09-068-377-1 x T43573

```

```

Align seg 1/1 to: T43573 from: 1 to: 1925

```

```

33 ArgLysMetCysLysAspValGluGluLeuLeuArgGlnAlaGlnAl 49
:::|||||::: :: |||:::|||||::: :::::||||
108 AAGAGCTCACCCCGCGCGAGAGAAAGCTTCTCCAAACCTGGGAGAC 157
49 aGluGluArgTyrGlyLysGluLeuValGlnIleAlaArgLysAlaGly 66
1::|::|::| :::::|||| :::::||||
158 AGATGAGACCAAGATGAGACAGTTTGACGACGTGCTCCAGAAATTCA 207
66 LysLp...ThrLp.....MetAsnSerLeuArgThrSerPhe 77
||||| |||||
208 AGCAGCTGACGAGAGGACCCGCGCTGACAGAGATCTCCGACCTTAC 257
|||||::|::: :: |||
258 GCTCTGCTCAAAAGCCATGCACGAG..... 281
94 uAlaLeuAlaLeuArgGluGluLeuArgSerLeuGluPheArgGluAl 111
|| ::::: :: |||:::||||
282 .....GCTTCCAAAGAGCTGAATGAGTGTCTGCAGAGGTGTATGAG 324
111 rGlnLysGluGluGlnArgLysTyrGluAlaIle..... 122
325 CCGATTGGCCCGCGAGGATGAGCAACACAGATGCGCAGAAACAGCAG 374
123 .....MetAspArgValGlnLysSerLysLeuSerLeuTyrLysL 136
375 CTGCTGTGATGATGATTACCAACAGAAAG.....CTGTGACACA 412
136 sThrMetLysLysLysAlaTyr.....AspGlnLysC 148
:::|::: ::::: |||
413 GCGCGCTGCTGACATGACACGTACTGTGGCCAGTTCCTCCGACATCA 462
148 ysArgAspAlaAspAspAlaGluGluGlnAlaPheGluArgValSerAla 164
||| ||| ::::: ::|
463 CACGCGATTGCCAAGCGGCGCAAGCTGTGTGCTACGACAGTGCCTGG 512
165 GlyHisGlnLysGlnValGlu.....LysSerGlnAsnLysAl 177
|| :::::|||||
513 CACCACTACGAGATCCCTTCAAACTGCCAAAGAAAGATGAAGCCAAAT 562
177 aLysGlnCysLysGluSerAlaThrGluAlaGluArgValTyrArgGln 193
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

563 TGCCAGGCGCGAGGAGGAGCTCATCAAGCCAGAGGTGTTTGAGGAGA 612
194 ..AsnIleGluInleuGluArgAlaArgThrGluTrpGluInGluHis 209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
613 TGAATGTGATCTGCAGAGGAGAGTGCCTGCTCCCTTGACAGACCGCGTA 662
210 ArgThrThrCysGluAlaPheGluLeu..... 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
663 GATTTCATGTCACACGCTCCAGAGCATCGGGCGCTGAGAGAAACTT 712
219 ....GlnIuPheAspArgLeuThrIleLeuArgSmaIleuTrpValH 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
713 CCACAGGAGATGAGCAAGCTCAACCAAGCTCATATGATGCTGCTG 762
224 IsCysAsnGln.....LeuSerMetGlnCys 242
|||||:|||||:|||||:|||||:|||||:|||||:
763 GCCTGCAAGACGACACGAGGACACCTTCACGGTCAAGGCCAGGCC 812
243 VallysAspAspGluLeuTrpGluGluValArgLeuThrLeuGluGlyCy 259
|||||:|||||:|||||:|||||:|||||:|||||:
813 AGAAAGAAAGTAAACTGTTTCCGCGCTGCGCAGAAAGAACAGTGA 862
259 sAspValGluGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyA 276
|||||:|||||:|||||:|||||:|||||:|||||:
863 CACGCGCGCTGCAAAAGGGAAC.....AAGAGCCCTTGGCTCAG 903
276 rgGluProAlaProValProTyrGlnAsnTyrTyrAspArgGluVal 292
|||||:|||||:|||||:|||||:|||||:|||||:
904 ATGGCTCCCTGCGCGCCACCC...GAGATCAGATCAACACAGAGCA 950
293 ThrProLeuIleGly.....SerProse 300
|||||:|||||:|||||:|||||:|||||:
951 GAGCGCGCGGGGGGCCACGCGCGGCCACCTCCCAAGTCCCAATC 1000
300 r.....IleGlnProSerCys 306
|||||:|||||:|||||:|||||:|||||:
1001 TCAGCCACACAGAGCTCGGAGCTGCGGGTGAGCCACCACTGCGGCTG 1050
306 lValIleLysArgPheSerGlyLeuLeuHisGlySerProLysThrThr 322
|||||:|||||:|||||:|||||:|||||:|||||:
1051 GAGCC.....CAGGAGCCAGGGGAGACTTCTCAAGTGAAGCAGCTCC 1094
323 ProSerAlaProAlaIleSerThrGluThrLeuThrProThrProGluAr 339
|||||:|||||:|||||:|||||:|||||:|||||:
1095 AGCTCTCTCTCTGCTGCTGTGTGAGACTTCCAGCAACT..... 1136
339 gAsnGluLeuValTyrAlaSerIleGluValGlnAlaThrGlnGlyAsnL 356
|||||:|||||:|||||:|||||:|||||:|||||:
1137 .....GTGAATGGCACCGTGAGGCGCGCAGTGGCGCGCGCT 1176
356 euAsn.....SerSerAlaGlnAspTyrArgAlaLeuTyrAspTyr 369
|||||:|||||:|||||:|||||:|||||:|||||:
1177 TGGACCTGCCCCCAGGTTTCATGTTCAGGTACAGGCCAGCAGACATAC 1226
370 ThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspIleLeuAl 386
|||||:|||||:|||||:|||||:|||||:|||||:
1227 ACGGCACTGACACAGACAGCTCAAGCTGCTGATGTGTGCT 1276
386 aValIle.....LeuGluGlyGluAspGlyTrp..... 395
|||||:|||||:|||||:|||||:|||||:|||||:
1277 GGTGATCCCTTCCGAACCTCGAAGACAGCAGATGAAGCTGCTCATAG 1326
396 .....TrpThrValGluArgAsnGlyGln.....Arg 404
|||||:|||||:|||||:|||||:|||||:|||||:
1327 GCGTGAAGAGAGAGCGATGGAACCGACAGAAAGACTGAGAAAGTCCGT 1376
405 GlyPheValProGlySerTyrLeuGluLysLeu 415
|||||:|||||:|||||:|||||:|||||:|||||:
1377 GGCCTCTTCCCGAGAACTTCAGTAGAGGCTC 1409
seq_name: N_Geneseq_36:V15692
seq_documentation_block:

```

```

ID V15692 standard; cDNA; 1925 BP.
AC V15692;
DE 01-JUN-1998 (first entry)
DE Human Binl cDNA.
KW Human; Box-dependent myc-interacting protein; Binl;
KW treatment; cancer; hyperplastic disease; neurodegenerative disease;
KW antibody; diagnosis; ds.
OS Homo sapiens.
FH Key location/Qualifiers
FT CDS 60..1415
FT FT /*tag= a
FT PD /product= Binl
PN US5723581-A.
PD 03-MAR-1998.
PF 24-MAY-1996; 652972.
PR 24-MAY-1996; US-652972.
PR 05-MAY-1995; US-435454.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Prendergast GC, Sakamuro D;
DR WPI; 98-178541/16.
DR P-PDB; M47295.
PT Box-dependent myc-interacting protein Binl - useful for, e.g.,
PT treating diseases associated with inadequate Binl levels such as
PT colon cancer.
PS Example 2; Columns 25-30; 32pp; English.
CC The present sequence encodes human Box-dependent
CC myc-interacting protein Binl, which can be used to treat diseases
CC associated with inadequate Binl levels, e.g. liver, colon or breast
CC cancer or hyperplastic diseases, or diseases where Binl activity
CC needs to be inhibited, e.g. neurodegenerative diseases, or to raise
CC antibodies for diagnostic purposes.
SQ Sequence 1925 BP; 467 A; 534 C; 564 G; 340 T;

```

alignment_scores:

```

Quality: 149.50 Length: 461
Ratio: 0.656 Gaps: 20
Percent Similarity: 49.458 Percent Identity: 21.692

```

alignment_block:

```
US-09-068-377-1 x V15692 ..
```

```
Align seg 1/1 to: V15692 from: 1 to: 1925
```

```

33 ArgLysMetCysLysAspValGluGluLeuLeuArgGlnArgAlaGlnAl 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
108 AAGAAGCTCACCCCGCGCGCAGAGAGAGTCTCCAGAGAGCTGGGAGAGC 157
49 gGluGluArgTyrGlyLysGluLeuValGlnIleAlaArgLysAlaGlyG 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
158 AGATGAGACCAAGATGATGAGAGTTTGAGCAGTGCCTCCAGATTTCAACA 207
66 lGln...ThrGlu.....MetAsnSerLeuArgThrSerPhe 77
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
208 AGCAGCTACGAGAGGCGCACCGGCTGCGAAGAGATCTCGGAGCTACCTG 257
78 AspSerLeuLysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLe 94
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
258 GCCTCGCTCAAGCCATGCACAGAG..... 281
94 uAlaLeuAlaLeuArgGluGluLeuArgSerLeuGluGluPheArgGluA 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
282 .....GCTTCCAAAGAGCTGAATGATGCTGCGAGAGGTGATAGGC 324
111 rgGluLysGluGluArgLysLysTyrGluAlaIle..... 122
|||||:|||||:|||||:|||||:|||||:|||||:
325 CCGATTGCGCGCGCAGGCGATGAGGCAACAGATCGCAGACACACAGAC 374
123 .....MetAspArgValGlnLysSerLysLeuSerLeuTyrLysLys 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
375 CTGCTGTGATGATGATTACACACAGAG.....CTGGTGAGACCA 412
136 sThrMetGlnSerLysLysAlaTyr.....AspGlnLysC 148

```

```

413 GGGCTGCTGACCATGACACAGCTGGCCAGTTCCCGACATCAAGT 462
148 ysaAspAlaAspAlaGluGlnAlaPheGluArgValSerAlaasn 164
463 CACCATTTGCCAAGCGGGGGCCAAAGCTGTGTGACTACGACAGTCCCGG 512
165 GlyHisGlnGluValGlu.....LysSerGlnAsnLysAl 177
513 CACCACTACGAGTCCCTTCAAACTGCCAAAAGAAAGATGAAGCCAAAT 562
177 alyGlnCysLysGluSerAlaThrGluAlaGluArgValIleArgGln 193
563 TGCCACAGCCGAGAGAGCTCATCAAGCCGACAGAGGTGTTGGAGAGA 612
194 .AsnIleGluGlnLeuGluArgAlaArgThrGluTrpGluGlnLys 209
613 TGAATGTGGATCTGCAGAGAGAGCTGCCGTCCCTGTGAAACAGCCGCTA 662
210 ArgThrThrCysGluAlaPheGlnLeu..... 218
663 GGTTCACGTCACACAGCTTCCAGAGCATCGGGGCTCGAGAAACTT 712
219 ...GlnGluPheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValH 234
713 CCACACAGAGATGAGCAAGCTCAACACAGAACCTCATGATGTGCTGTCG 762
234 IscysAsnGln.....LeuSerMetGlnLys 242
763 GCCTGGAGAACCAACACGGGACCAACCTTCAAGGTAAAGGCCAGCC 812
243 ValLysAspArgLeuLeuTrpGluValAlaArgLeuThrLeuGluGlyC 259
813 AGAAGAAAGAAAGTAACTGTTTTCGGGCTGCGCAAGAAAGAACAGCA 862
259 sasPvalGluGlyAspIleasnGlyPheIleGlnSerLysSerThrGlyA 276
863 CAACGCCCTGCAGAAAGGAGAAC.....AAGAGCCCTTCGCTCCAG 903
276 rgLuPProPAlaProValProThrGlnAsnTrpTyrAspArgLysVal 292
904 ATGGCTCCCTCCGCCGCCACCC...GAGATCAAGATCAACACAGACCA 950
293 ThrProLeuIleGly.....SerProSe 300
951 GAGCCGCGCGGGGGCCAGCCGCGGGCCACCTCCCGCCAGTCCCAATC 1000
300 F.....IleGlnProSerCysG 306
1001 TCAGCCAGCAGAGGCTTCGAGGTGGCGGTGGAGACCAACCTCGCGCTG 1050
306 lyValIleLysArgPheSerGlyLeuLeuHisGlySerProLysThrThr 322
1051 GAGCC.....CAGAGCCAGGGGGAAGCGCGCAAGTGAAGACGCCCTC 1094
323 ProSerAlaProAlaAlaSerThrGluThrLeuThrProThrProGluVal 339
1095 AGCTCTCTCTGCTGTCGTGTGTGAGAGACCTTCCAGCAACT..... 1136
339 gaSnGluLeuValTyrAlaSerIleGluValGlnAlaThrGlnGlyAsnL 356
1137 .....GTGAAATGGACACCTGGAGGGGCGCAGTGGGGCGGGCGCT 1176
356 euasn.....SerSerAlaGlnAspTyrArgAlaLeuTyrAspThr 369
1177 TGAACCTGCCCCCAAGTTTCATGTTCAAGGTACAGGCCCGCCAGCAGCTAC 1226
370 ThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspIleLeuAl 386
1227 ACGGCCACCTGCACACAGACGAGCTCAGCTCGTGTGATGTGTGCT 1276
386 aValIle.....LeuGluGlyGluAspGlyTrp..... 395
|||||
|||||

```

```

1277 GGTGATCCCTTCACAGAACCTGAGAGCAGATGAGGCTGGCTCATGG 1326
396 .....TrpThrValGluArgAsnGlyGln.....Arg 404
1327 GCGTGAGAGAGACCGACCTGGAACACACACAGAGCTGAGAAAGTCCCT 1376
405 GlyPheValProGlySerTyrLeuGluLysLeu 415
1377 GCGCTCTTCCCGACAGACTTCACTGAGAGGTC 1409
seq_name: N_Geneseq_36: X16322

```

```

seq_documentation_block:
ID X16322 standard; cDNA; 1925 BP.
AC X16322;
DT 22-APR-1999 (first entry)
DE Human Bin1 protein encoding cDNA.
KW Bin1; brain-specific; box-dependent myc-interacting protein; cancer;
diagnosis; hyperplastic disease; tumour suppressor; gene therapy;
KW benign prostatic hypertrophy; neurodegeneration; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 60..1415
FT CD5 /*tag= a
FT WO9855151-A1.
PN 10-DEC-1998.
PD 04-JUN-1998: U11647.
PR 06-JUN-1997; US-870126.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Prendergast GC, Sakamuro D;
DR P-PSDB: W94504.

```

```

PT New splice variant exons of box-dependent myc-interacting
PT polypeptide - associated with loss of tumour suppressor activity,
PT used for diagnosis of cancer and in gene therapy
PS Example 1; Page 78-81; 133pp; English.
CC The present invention describes Bin1 (Box-dependent myc-interacting
CC protein 1) brain-specific alternative splice variants exon 12A, B, C and
CC D. Bin1 specific antibodies (when labelled) are used to detect cancers
CC or other hyperplastic conditions (e.g. benign prostatic hypertrophy)
CC associated with a deficit of normal Bin1 and/or aberrant forms of Bin1.
CC The same diseases can also be diagnosed at the nucleic acid level using
CC fragments of Bin1 nucleotide sequences in standard amplification and/or
CC hybridisation assays. Cancers that can be detected are carcinomas and
CC epithelial cell tumours, specifically of prostate, liver and colon/
CC rectum, also melanoma. Bin1 nucleotide sequences, when included in
CC a vector, and Bin1 proteins may also be used to treat these diseases,
CC and also degenerative conditions such as neurodegeneration. Bin1
CC proteins may also be used to raise antibodies, and the nucleotide
CC sequences can be used to express the corresponding proteins. Also
CC anti-idiotype antibodies can be used therapeutically, and more
CC generally similar materials can be used to treat or diagnose any
CC condition involving deregulation, defect or amplification of the c-myc
CC oncogene. Bin1 nucleotide sequence, proteins and antibodies are also
CC useful to screen for agents that may be used to treat Bin1-related
CC cancers. The present sequence encodes human Bin1 protein.
SQ Sequence 1925 BP; 467 A; 534 C; 584 G; 340 T;

```

```

alignment_scores:
Quality: 149.50 Length: 461
Ratio: 0.656 Gaps: 20
Percent Similarity: 49.458 Percent Identity: 21.692

```

alignment_block:

```

US-09-068-377-1 x X16322 ..
Align seg 1/1 to: X16322 from: 1 to: 1925

```

```

33 ArgLysMetCysLysAspValGluGluLeuLeuArgGlnArgAlaGlnAl 49
|||||
|||||
108 AAGAGCTCACCGCGCGCAGAGAAAGTTCTCCAGAGCTGGGGAAAGCC 157
|||||

```


CC erythematous, multiple sclerosis, myasthenia gravis, osteoarthritis,
 CC osteoporosis, pancreatitis, polycystic kidney disease, polymyositis,
 CC scleroderma, Sjogren's syndrome, autoimmune thyroiditis, complications
 CC of cancer, extracorporeal circulation, viral, bacterial, fungal,
 CC parasitic, protozoal and helminthic infections, and trauma. Antibodies
 CC specific for IRS-p53h may be used directly as an antagonist or indirectly
 CC as a targeting or delivery mechanism for bringing a pharmaceutical
 CC agents to cells or tissue which express IRS-p53h.
 SQ Sequence 2080 BP; 440 A; 657 C; 622 G; 360 T;

alignment_scores:

Quality: 148.50 Length: 439
 Ratio: 0.678 Gaps: 21
 Percent Similarity: 49.886 Percent Identity: 21.640

alignment_block:

US-09-068-377-1 x V72220 ..

Align seg 1/1 to: V72220 from: 1 to: 2080

```

35 MetcyslyspvalglugluLeuValArgGlnArgGlnAlaGlnGlu 51
195 ATGGGGAAGAAATACAGAGAGACACTGGACGTGTGACATGACCCAA 244
51 uArgTyrGlyLysGluLeuVal.....GlnIleAlaArgLysAlaG 65
245 AGGCTACTTTCACGCCCTGTGTGAAGATGGGGAGCTGGCCAGCAGACC 294
65 LylGlyGlnThrGlnMetAsnSerLeuArgThrSerPheAspSerLeuLys 81
295 AGGCTCCCAAGAACTCGAGACGTTCTTCAGATGCTGAAATGTCAC 344
82 GlnGlnThrGluAsn.....ValGlySerAlaHisIleGlu 93
345 AGGAGATCCAGAAATCAGCTGGAAGAAATGCTTTCACAGACGA 394
93 nLeuAlaLeuAlaLeuArgGluLeuArgSerLeuGluLysPhe... 108
395 GCTGCTTACCCAGCTGAGAGAGAGAGTGTGACATCCAGATTCGA 444
109 .....ArgGluArgGlnLysGluLysArgLysLysTyrGluAla 121
445 GTGCTGGCTAAAGAAATACCAAGTCTGAGCAAGAGCAAGCAGCC 494
122 IleMetAspArgValGlnLysSerLysSerLeuTyrLysLysThrMe 138
495 ...CTGGCAAGTGTCAAGCTGAGCAAGAGCTTCGAAAGAGAGCA 541
138 tGluSerLys.....LysAlaTyrAspGlnLysCysArgAspAla 152
542 GCGCAGCAGAAATCTCAGAGTACTCGCAGACAGAGCTGCAGTACATCG 591
152 sp.....AspAlaGluGlnAlaPheGluArgValSerAlaAsnGly 165
592 ACGCATCAGCAACAGCAGGCGAGCTGAGAAATTAAGTGTCCGACGCG 641
166 HisGlnLysGlnValGluLysSerGlnAsnLys.....Al 177
642 TACAGACCCGACACTGACAGAGAGAGGCGCTTCTGCTTCCTGTGGA 691
177 aLysGlnCysLysGlnSerAlaThrGluAlaGluLysValTyrArgGln 194
692 GAACACAGTGGCCGCTGGCAGAACTCCGCG...GCCATCAGACTCCAAAG 738
194 snIleGluGlnLeuGluArgAlaArgThrGluTyrPheGluGlnHisArg 210
739 GCAAGGAGCTGCTGCGCAGAGCTGCGCTGTGCAACAG..... 779
211 ThrThrCysGluAlaPheGlnLeuGlnLysPheAspArgLeuThrIle 227
780 ...GCTGTGCG.....GACCCAGCAGAGATCC 805

```

```

227 uArgAsnAlaLeuTrpValHisCysAsnGlnLeuSerMetGlnCysVal 244
806 GGAGCCGCGGTTG.....CAGCTCATGACAGCAGTGGGCA 840
244 yAspAspLeuLeuTyrGluGluValArgLeuThrLeuGluGlnCysAsp 260
841 GCAMAGGC.....GCCACCTCCACAGCCGCTG 869
261 ValGluGlyAspIleAsnGlyPheIleGlnSerLysSerThrLysArg 277
870 TCGGCTCCAGTCCACCTGCTATTCGACCCCATTCGGGGGCA 919
277 uProAlaProValProTyrGlnAsnTyrTyrAspArgGluValThrP 294
920 G...CCCTGCGGTTGCC.....CCGACCTGGCAC 948
294 roLeuIleGlySerProSerIleGlnProSerCysGlyValIleLysArg 310
949 CGTTCGTGGGGGATGCTGCCAGAGAGACACCATCATGAACGCG 998
311 PheSerGlyLeuLeuHisGlySerProLysThrThrProSerAlaPro.. 326
999 GTCCACAGGC.....CCGATGCGAGAGACTACAGCCCGTG 1033
327 .....AlaAlaSerThrGluThrLeuThrProThrProGlu 1339
1034 GCGTGACCCGCAAGCGCTGCGCCAGCCAAATCCCTCTCTCCCGCAGTTC 1083
339 rGAsnGluLeu.....ValTyrAlaSerIleGluValGlnAlaThrGln 354
1084 AAGCAGAGCTCAGCGACTCTTACTCCACACACTCCCGTGCGCAAGAGC 1133
343 .....ValTyrAlaSerIleGluValGlnAlaThrGln 354
1134 GTGACCCCAAAAAACAGCTATGCTACACCGAGAACAGACTGCTGCTG 1183
354 yAsnLeuAsnSerSerAla.....GlnAspTyrA 364
1184 CTCGACCTCATGCGACGCGCGCTGAGCGCATGGCCGATGCGGGTGA 1233
364 rGAlaLeuTyrAspTyrThrAla..GlnAsnSerAspGluLeuAspIle 379
1234 AGGCACTCTTCTCCCGCTGCGGAGACACACCTCTCTGAGACTTC 1283
380 SerAlaGlyAspIleLeuAlaValIleLeu..GluGlyGluAspLys 395
1284 AAGGAGGTGACCTCATTAACCTGCTGCTGCTGAGGCGCGCATGGCTG 1333
395 P.....TrpThrValGluArgAsnGlyGlnArgGlyPheValProGly 410
1334 GCACATCAGAGAGAGTGAAGAACCAAGATGCGGCGCTGTTCCCTTC 1383
410 eTTrLeuGluLysLeu 415
1384 CCTACACCGGCGCTTG 1400

```

seq_name: N_Geneseq_36:139799

seq_documentation_block:

ID T39799 standard; DNA; 2873 BP.

AC T39799;

DT 19-FEB-1998 (first entry)

DE Human clone 65 gene.

KW Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;

KW cellular signalling element; cellular structural element; malignancy;

KW protein identification; functional domain; protein screening;

KW cellular signal transduction process; ss.

OS Homo sapiens.

OS key Location/Qualifiers

FT CDS 3..1532

FT /**tag= a

FT PD WO9631625-A1.

PD 10-OCT-1996.

```
alignment_block:
US-09-068-377-1 x T39799 .
```

```

126 valGlnIysSerLysLysSerLeuIYrLysLysIstMMeGlnSerLysLy 14
15 ATTGAAAGGAAAGAAATGATTAGACCTA.....ATCCGAGAAAAAGAA 52
142 salATyAspGlnIlnLysCysArGyAspAlaAspAspAlaGlnIlnAlaPheG 156
53 ACTGGAAGATGAGAGCTGCACAG.....AAACCAAGACAGAGAAAG 93
159 lUaArGValSerAlaAsnGlyHsGlnLysGlnValGlnLysSerGlnAsn 175
94 AAACCTATATGGAAGAAATCTTACAAAGGAGAGAAAGAAACAAACAAAG 143
176 LysAlaLysGlnCysLys.....GluseralatHrGt 186
144 CGAGCTCCAGGAAACAAAACACAAACAAAATTCACAGAGAGGACGGAA 193
186 uAlaGlnUArGValTYArGlnAsnIleGlnIlnLeuGlnUArGAlaArgT 203
194 AGCTGAGGAGAAACAACAGCTGACACAGCTAGAGTTTGGTAATTATAGAG 243
203 hUcIuTrpGlnGlnIlnHsArgThrHrThrcysGlu.....Ala 215
244 CATTATATCCCTCTTGAACCAAGAAACCATGATGATGATGATTTAAATTGT 293
216 PheGlnLeuGlnGlnUArPheAspArgLeuThrIle..... 226

```

```

seq_name: N_Geneseq_36.q054841
seq_documentation block:
ID 054841: standard; CDNA; 6306 BP.
AC 054841:
DF 15-JUL-1994 (first entry)
DE Sequence encoding inner nuclear matrix (INM) protein MIM2
KM Malignant cell type marker; interior nuclear matrix; MIM2;
OS Homo sapiens.
FH Key
FT cds
FT 1. 6306
   Location/Qualifiers
   /tag= a
PN MO9400573-A.
PD 06-JAN-1994.
PE 21-JUN-1993; U06160.
PR 22-JUN-1992; US-901701.
PA (MATR-) MATRITTECH INC.
PI Lidgerd GP, Youkaly G;
DR WPI; 94-026210/03.
DR P-PSDB; R47143.
PT Novel malignant cell type markers of the interior nuclear matrix
PT - used for detecting abnormal cell types e.g. malignant breast,
PT prostate, lung, etc., for determining deg. of cell death in

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2000, 09:15:45 ; Search time 59.02 Seconds
(without alignments)
4894.396 Million cell updates/sec

Title: US-09-068-377-2

Perfect score: 2100
Sequence: 1 caatattcaagctatacca.....gaaggaaaaaaaaaaaaaa 2100

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_MA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PTMUS.COMB.seq:*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2100	100.0	2100	5 US-08-938-830-2	Sequence 2, Appli
2	803.4	38.3	1613	5 US-08-938-830-28	Sequence 28, Appli
3	343.8	16.4	980	2 US-08-452-267-1	Sequence 1, Appli
4	111	5.3	231	6 PCT-US95-06722-20	Sequence 20, Appli
5	78.8	3.8	7218	1 US-08-232-463-14	Sequence 14, Appli
6	63.4	3.0	5452	4 US-09-130-114-1	Sequence 1, Appli
7	63.4	3.0	10596	1 US-07-885-971-15	Sequence 15, Appli
8	63.4	3.0	10596	1 US-08-087-783A-15	Sequence 15, Appli
9	63.4	3.0	10596	1 US-08-194-088B-15	Sequence 15, Appli
10	63.4	3.0	10596	1 US-08-194-087-15	Sequence 15, Appli
11	63.4	3.0	10596	6 PCT-US93-04648-15	Sequence 15, Appli
12	63.4	2.9	3489	4 US-08-728-323A-1	Sequence 1, Appli
13	61.6	2.9	32207	3 US-08-770-379-20	Sequence 20, Appli
14	61.2	2.9	1931	4 US-09-130-114-2	Sequence 2, Appli
15	50	2.4	1137	7 5171843-8	Sequence 1, Appli
16	48.8	2.3	2338	4 US-08-425-069-1	Sequence 1, Appli
17	48.8	2.3	2338	4 US-08-317-844B-1	Sequence 1, Appli
18	47.8	2.3	2887	3 US-08-533-306A-3	Sequence 3, Appli
19	47.8	2.3	2887	3 US-08-742-923A-3	Sequence 3, Appli
20	47.2	2.2	1430	4 US-08-549-004A-15	Sequence 15, Appli
21	46.8	2.2	697	7 5171843-10	Sequence 3, Appli
22	46.6	2.2	1995	2 US-08-425-069-3	Sequence 3, Appli
23	46.6	2.2	1995	2 US-08-317-844B-3	Sequence 3, Appli
24	46.6	2.2	2680	3 US-08-533-306A-5	Sequence 5, Appli
25	46.6	2.2	2680	3 US-08-742-923A-5	Sequence 5, Appli
26	46.6	2.2	2680	3 US-08-742-923A-5	Sequence 5, Appli

27	45.6	2.2	1045	1 US-08-475-894-5	Sequence 5, Appli
28	45.6	2.2	1045	1 US-08-484-710-5	Sequence 5, Appli
29	45.6	2.2	1045	3 US-08-484-709-5	Sequence 3, Appli
30	45.2	2.2	1392	1 US-08-475-894-3	Sequence 3, Appli
31	45.2	2.2	1392	1 US-08-484-710-3	Sequence 3, Appli
32	45.2	2.2	1392	3 US-08-484-709-3	Sequence 3, Appli
33	45.2	2.2	1659	1 US-08-475-894-1	Sequence 1, Appli
34	45.2	2.2	1659	1 US-08-484-710-1	Sequence 1, Appli
35	45.2	2.2	1659	1 US-08-484-709-1	Sequence 1, Appli
36	43.8	2.1	2003	3 US-08-942-423-1	Sequence 1, Appli
37	42.6	2.0	3134	3 US-08-533-669A-1	Sequence 1, Appli
38	42.2	2.0	1003	1 US-07-800-364B-13	Sequence 13, Appli
39	42.2	2.0	1003	3 US-07-989-847-11	Sequence 11, Appli
40	42.2	2.0	1723	1 US-07-841-646-28	Sequence 28, Appli
41	42.2	2.0	1723	1 US-07-901-703-10	Sequence 10, Appli
42	42.2	2.0	1723	1 US-08-147-023-28	Sequence 28, Appli
43	42.2	2.0	1723	1 US-08-206-864-20	Sequence 3, Appli
44	42.2	2.0	1723	1 US-08-278-729A-20	Sequence 20, Appli
45	42.2	2.0	1723	6 PCT-US92-01968-20	Sequence 20, Appli

ALIGNMENTS

```

RESULT 1
US-08-938-830-2
: Sequence 2, Application US/08938830
: Patent No. 6040437
:
: GENERAL INFORMATION:
: APPLICANT: Lasky, Laurence A.
: ATTORNEY/AGENT INFORMATION:
: DREGER, GINGER R.
: FILING DATE: 07-FEB-1997
: PRIORITY APPLICATION DATA:
: PRIOR APPLICATION NUMBER: 08/798419
: FILING DATE: 07-FEB-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Dreger, Ginger R.
: REGISTRATION NUMBER: 33,055
: REFERENCE/DOCKET NUMBER: P1066P1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650/225-3216
: TELEFAX: 650/952-9881
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2100 base pairs
: TYPE: Nucleic Acid
: STRANDEDNESS: Single
: TOPOLOGY: Linear
:
: US-08-938-830-2

```

Query Match 100.0%; Score 2100; DB 5; Length 2100;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caatlattcaagctataccaagcatatacaatccaagtctacccaagaagacg 60
 Db 1 CAATATTTCAGGCTATACCAAGCATATCACTCAAGCTTATGCCCAAGAAGACG 60
 QY 61 gaaggtctgagcgcccaatttaacaaagtggaaatctgtcgtatgtcattgtc 120
 Db 61 GAAGGTCTCAGGCGGCCCAATTTTAATCAAAAGTGGAAATATGCTATGCTATTGTC 120
 QY 121 ctccacttacttaacagttagcaacggtccgaacctatacaactcaaaacttca 180
 Db 121 CTTCACCTTTACTATACAGTATGCAACAGSTCCGAACTCATATCAACTCAAACTATTCTCA 180
 QY 181 agcgcttcaacccaatttgcctcctccttaacgltcaatgataacttcaatgaat 240
 Db 181 AGCGCTTTCACCAACCAATTTGCTCTTCAAGTTCAATGATTAATTAATGAAT 240
 QY 241 caaggtcagtaaaatttgaatgttaataatcaaaacacgttcaactgttgaagga 300
 Db 241 CAGGCTAGTAAATGATATGATTAATTAATTAATTAATTAATTAATTAATTAAT 300
 QY 301 ccaactgtctataacgcttggaaactacacagagatgtttaataccaactacacg 360
 Db 301 CCAACTGCTATTAACGCTTTTGGAACTACTACAGGATGTTTAAATCCACTACAAATGGA 360
 QY 361 tgaatataactatcaltcgaatgaatgaatgaatgaatgaatgaatgaatgaat 420
 Db 361 TGATGTATATATCTATCTATTCATGATGAATGATGAATGATGAATGATGAAT 420
 QY 421 tgggtgacccaacgctcggcgtcctcctcctcctcctcctcctcctcctcctc 480
 Db 421 TGGGTGACCCACGCGCTCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
 QY 481 acaaggttgcctcttctcctcctcctcctcctcctcctcctcctcctcctcct 540
 Db 481 AAGAGTGTAGGCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
 QY 541 cgttcaacgagcgccgcttggagtgaggagagagagagagagagagagagagag 600
 Db 541 CGTTTACAGAGCGCGCGCTGGAGAGTGGAGAGCGCGCTGGAGAGCGCGCTGGAGAG 600
 QY 601 ggaagctgtcctcgtcctcgtcctcgtcctcgtcctcgtcctcgtcctcgtcct 660
 Db 601 GGAAGCTGTCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 660
 QY 661 ttgacaacagcctctgagagatgagtgagtgagtgagtgagtgagtgagtgag 720
 Db 661 TTGACACAACGCGCTTGAGAGATGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAG 720
 QY 721 aggaacttcaagcccaacaaaggtatgagtgagtgagtgagtgagtgagtgag 780
 Db 721 AGGAGCTTTCACGCGCCACACAGAGGTATGAGTGAGTGAGTGAGTGAGTGAGTGAG 780
 QY 781 aagaatgcaagagatgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 840
 Db 781 AAGATGTGCAAGAGATGTGAGAGAGTGCTCTGAGACAGAGAGAGAGAGAGAGAG 840
 QY 841 gggaaagagcctgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 900
 Db 841 GGAAGAGAGCTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAG 900
 QY 901 aggaactccttgaactcctgaaagcaagcaagcaagcaagcaagcaagcaagcaag 960
 Db 901 AGGAGCTCTCTTTCACCTCTGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 960
 QY 961 ctgagcctgagcctgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1020
 Db 961 CTGAGCCTGAGCCTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAG 1020
 QY 1021 gaggcaggaagaagatagagagcatatgagagcgttgcagaagaagcaagtgctgc 1080
 Db 1021 GAGCAGCGGAAGAAGATAGAGGCAATCATGAGCAGCTGTGCCAAGAAGCAAGTGCTGCTC 1080
 QY 1081 tacaagaagacatgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1140

Db 1081 TACAAGAAGACATGAGTCCAGAAAGCATATGACCAGAAGATGACAGATGAT 1140
 QY 1141 gctgagagagcctctgagcgttggatgtgcaatgagcccaagaagcagatagaagaagc 1200
 Db 1141 GCTGAGAGAGCCTTTCAGCGCTGTGAGTGCCAAATGGCCACAGAAAGTGAAGAAAGAGC 1200
 QY 1201 cagaacaaagccaagcagtgcaaggaatcagccaagaagcagaagaagtgtcaagcaa 1260
 Db 1201 CAGACAAACCAAGCAGTGTCAGAGACTCAGCCACAGAGGACAGAAAGTGTACAGGCA 1260
 QY 1261 aatatgaacaaactggaagagcaggaacccagtgaggagcaaggaacagcagactact 1320
 Db 1261 AATATGAAACAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 QY 1321 gaggcctcagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1380
 Db 1321 GAGGCTTTCAGTTGCGAGGAGTTGAGCGGCTCACATCTCTCCGCAATGCCCTGTGGTG 1380
 QY 1381 cactgtaacagctctcactgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1440
 Db 1381 CACTGTATACAGCTCTCTCATGAGTGTGTCAAGATGATGATGATGATGATGATGATGAT 1440
 QY 1441 ctgacccctgagagcgttgaatgagtgagtgagtgagtgagtgagtgagtgagtgag 1500
 Db 1441 CTGACCCCTGAGGAGCTGTGATGAGTGTGATGATGATGATGATGATGATGATGATGAT 1500
 QY 1501 actgagagaagccccaagctcgggtgagtgagtgagtgagtgagtgagtgagtgagtgag 1560
 Db 1501 ACTGAGAGAGCCCGGAGCTCGGAGTCTTATGAGTGTGATGATGATGATGATGATGATGAT 1560
 QY 1561 ccaactgagtgagcctcactgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1620
 Db 1561 CCACTGATTTGAGAGCCTTACATGACAGCCCTCTCGGAGTGTGATTAAGAGAGTCTCTGGG 1620
 QY 1621 ctgctacatgagagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1680
 Db 1621 CTGCTCATGTGAAGTCCCAAGACACACCTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 1680
 QY 1681 actcccaacccctgagcgaatgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1740
 Db 1681 ACTCCCAACCCCTGAGCGGAATGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 1740
 QY 1741 ggaacacttaactcaatgaagcccaagagtaacagcagcactcactacactacagat 1800
 Db 1741 GGAACCTTAACATCAATCAGCCCAAGAGTACCGGCACTTACAGCAATACACTGACACAGAT 1800
 QY 1801 tctgagagcttgagacatttcgcgagagagatacctcgtcgttcaatcctcgtgagagagat 1860
 Db 1801 TCTGATGAGCTGAGACATTTCCGCGGAGACATCTGCGGCTCATCTCGAAGGGAGAGAT 1860
 QY 1861 ggtctgtgagccttggagcgaagcgaacgcttgccttgcctgagtgagtgagtgag 1920
 Db 1861 GGCTGTGAGACTGTGAGCGGAACGGAACGAGTGTGCTTGTCTCTGCTGCTGCTGCTGAG 1920
 QY 1921 aagccttgaggaagcctgagcagtlccacatacctcgcctgactgtgagtgagtgag 1980
 Db 1921 AAGCTTGAGGAAGCCTTACAGTCTTCCATACATCTCCGCTGAGTGTGAGTGAGTGAG 1980
 QY 1981 tgttcttctcactcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcag 2040
 Db 1981 TGTTCCTTTCATCACCGCCAGGCTCTCAGGCGGCGAGAAACGCGGCTGTGCTGGG 2040
 QY 2041 catggcctgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 2100
 Db 2041 CATGGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100

RESULT 2
 US-08-938-830-28
 ; Sequence 28, Application US/08938830
 ; Patent No. 6040437
 ; GENERAL INFORMATION:


```
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZypc-F15
US-08-232-463-14
```

```
Query Match 3.8%; Score 78.8; DB 1; Length 7218;
Best Local Similarity 5.0%; Pred. No. 2,7e-12;
Matches 20; Conservative 241; Mismatches 143; Indels 0; Gaps 0;
```

```
QY 935 agaatgtggagcgtcacacatccatcagcttgccctcgtgagagctgcgagcc 994
||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
DB 1446 AGATTGTGTCRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1387

QY 995 ttgagagcttcgagagacagaaagacagcaggaagatgagccatcagcc 1054
: : : : : : : : : : : : : : : : : : : : : : : :
DB 1386 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1327

QY 1055 gtgcacgaagagaagtgtcgctctacaagaagacatgagtcacaagaagc 1114
: : : : : : : : : : : : : : : : : : : : : : : :
DB 1326 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1267

QY 1115 accgaagtcgagagatgcatgatctgagcagccttcgaagctgtgagtc 1174
: : : : : : : : : : : : : : : : : : : : : : : :
DB 1266 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1207

QY 1175 gccacgaagcagtgaaagagcagaacaaagcgaagcagtcgaagagtcag 1234
: : : : : : : : : : : : : : : : : : : : : : : :
DB 1206 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1147

QY 1235 cagagcagaagaagtgtacgcaaatatcgaacaactggaagagcagagcc 1294
: : : : : : : : : : : : : : : : : : : : : : : :
DB 1146 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1087

QY 1295 gggagcagagcagcagctcgtgagccttcagtcag 1338
: : : : : : : : : : : : : : : : : : : : : : : :
DB 1086 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1043
```

```
RESULT 6
US-09-130-114-1/c
; Sequence 1, Application us/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: HOTLICK, Robert A.
```

```
APPLICANT: DamaJ, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; FILE REFERENCE: 0867/1D903U51
; CURRENT APPLICATION NUMBER: US/09/130,114
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: YEBNA
US-09-130-114-1
```

```
Query Match 3.0%; Score 63.4; DB 4; Length 5452;
Best Local Similarity 43.7%; Pred. No. 5.2e-08;
Matches 329; Conservative 0; Mismatches 421; Indels 3; Gaps 1;
```

```
QY 701 ccgagatgccttcgtgcagggatccacggccacacagggatgagtgctactgc 760
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2194 TCCAAGTTCATGTGCTCCAAAGGACCACGGTGGACAGAGCAGAGAGCGGG 2135

QY 761 agagctgtgacgycaggaagatgtgcaagatgtgagagctgtctcagacag 820
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2134 AGGGCAGAGCAGAGAGGGGCGAGGAGCAGAGAGAGGGGCGAGGAGGAGGG 2075

QY 821 cccagcgaggaagagtgtaagcggagagctgtgtgcaattgcagcagagctgtg 880
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2074 AGGGCAGAGAGGGGCGAGGAGGGCGAGGAGCAGAGAGCGGCGAGAGCAG 2015

QY 881 agacagagatgaattcccttgaggacctcttctgacctcctgaagagaacaag 940
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2014 AGGAGGGCAGAGAGGGGCGAGGAGCGAGGAGGGCGAGGAGCAGAGAGGG 1955

QY 941 tggcagtcacacatccagcttgccctgtgctgtgtgagagctgtgagcct 1000
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1954 GCGAGAGCAGAGAGAGGGGCGAGGAGGGGCGAGGAGGGCGAGGAGGAGGG 1895

QY 1001 agttccgagaagacagaagaagcagcggaaagtatgagccatcagacgtgt 1060
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1894 AGCAGAGAGGGGCGAGGAGGGGCGAGGAGCAGAGAGAGGGGCGAGGAG 1835

QY 1061 aagaagaagctgtcgctctcaagaagacacatgagatccaaagggcata 1120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1834 AGGAGCAGAGAGAGGGGCGAGGAGCGAGGAGGAGGAGGAGGAGGAGGAG 1775

QY 1121 agtgcagagatgcatgtgtgagcagccttcgaagctgtgtgagtcgaat 1180
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1774 AGGGCAGAGAGAGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGAGGAGG 1715

QY 1181 aagaagaagtgaagaagcagaacaaagcgaagtgcaagagtcagccagag 1240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1714 AGCAGAGAGGGGCGAGGAGGAGGAGGGGCGAGGAGGAGGAGGAGGAGG 1658

QY 1241 cagaagaagtgtacgcaaatatcgaacaacttgaagaagcagagcgagtg 1300
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1657 AGGAGGGCAGAGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1598

QY 1301 agagcagcagactcgtgtgagccttcagctgtgagagatlttgaacggct 1360
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1597 AGGAGGGCAGAGAGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGAGGAGG 1538

QY 1361 tcgcaatgacctgtggtgtgcaactgttaaccagcttcctcagtgagtg 1420
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1537 AGGAGCAGAGAGGGGCGAGGAGGCGAGGAGCAGAGAGGGGCGAGGAGG 1478

QY 1421 agctctatgaggaagtgcgtgagaccttgag 1453
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1477 AGGGCAGAGAGCAGAGAGGGGCGAGGAGGAGGAGG 1445
```

RESULT 7
US-07-884-811-15
; Sequence 15, Application US/07884811
; Patent No. 5316921
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/884,811
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 755,1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-07-884-811-15

Query Match 3.0%; Score 63.4; DB 1; Length 10596;
Best Local Similarity 43.7%; Pred. No. 7 6e-08;
Matches 329; Conservative 0; Mismatches 421; Indels 3; Gaps 1;
QY 701 tccgagatgcctctctgtgacaggaactcagccacacagaaggtatgaagtactatgc 760
DB 2191 tccagtttcatttgctgcaaaaggagccacggttgacacagagcagagacgagcggg 2250
QY 761 agaggtctgtgacgaggaagatgtgcaaggaatgtggaagagctgcctcagaacaggg 820
DB 2251 AGGGGACAGCAGCAGGAGGAGCAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2310
QY 821 cccagcgcgaagagagatgcggaagagagctgtgcaagattgcagcaagagctcgtgacc 880
DB 2311 AGGGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2370
QY 881 agacaagaatgaattcccttgaagacctcttgaactcccttgaagcagcaacaagaatg 940
DB 2371 AGGAGGGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2430
QY 941 tggagcagtgacacatcccaagctgtgcccctgtgagggagctgagcagccttgagg 1000
DB 2431 GGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2490
QY 1001 agtccgagagagagagaagacagcagcgaagaagatgaagagcaltgaaccgtgtcc 1060
DB 2491 AGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2550
QY 1061 agaagagcaagtgtcgtctctacagaagaacaltggaagttccaaagaagcatatgaccaga 1120

DB 2551 AGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2610
QY 1121 agtgcagagatgcagatgatgtcgtgacagagccttcagagctgtgagtgccaatgcccacc 1180
DB 2611 AGGGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2670
QY 1181 agaagcaagtagaagaagcagacaaagcgaagcagtgcaagagatgcaccagaagg 1240
DB 2671 AGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2727
QY 1241 cagaagaagtgtacagcaaaaatcgaacaactggaagagacgaagacagctgaggc 1300
DB 2728 AGGAGGGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2787
QY 1301 agagcaccagactacactgtgaagccttcagtgcaagagttgaccggtccaccatcc 1360
DB 2788 AGGAGGGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2847
QY 1361 tccgcaatgcctgtgtgactgttaaccagctctccatgcagtgtgtcaaggaatgatg 1420
DB 2848 AGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2907
QY 1421 agctctatgaggaagtcgagctgacccttgagg 1453
DB 2908 AGGGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2940

RESULT 8
US-07-885-971-15
; Sequence 15, Application US/07885971
; Patent No. 5328837
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,971
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-07-885-971-15

Query Match 3.0%; Score 63.4; DB 1; Length 10596;

TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-194-087-15

Query Match 3.0%; Score 63.4; DB 3; Length 10596;
Best Local Similarity 43.7%; Pred. No. 7.6e-08;
Matches 329; Conservative 0; Mismatches 421; Indels 3; Gaps 1;

QY 701 tccgagatgctcttggtgacggaacttcacgcccacacaggtatgagtgtactgc 760
DB 2191 TCCAGATTGCAATTGGCTGCCAAGGGACCCACGGTGGAAACAGGAGCAGAGCAGCGG 2250
QY 761 agaggtctctgagcagcaggaagatgtcaagatgtgaggaagctgtctcagacagagg 820
DB 2251 AGGGGACAGAGAGGAGGAGGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2310
QY 821 cccagcggaggaagaggtacggaagagctgtgacagtgcagcgaagctgtgtgc 880
DB 2311 AGGGGACAGAGGAGGAGGAGGAGGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGG 2370
QY 881 agacagagatgaattcccttgaggaacctcttgaacctcctgaagcagaacagagaatg 940
DB 2371 AGGAGGGGACAGAGGAGGAGGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2430
QY 941 tggagcagtcacacatccagctggccctggccctgctgtgaggaagctgtgagccctggagg 1000
DB 2431 GGCAGGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2490
QY 1001 agttccgagagagacagaagaagcagcggaaagatgtatgaggccatctgagccgtgtcc 1060
DB 2491 AGCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGG 2550
QY 1061 agaagagcaagttgtcgtctctacaagaagaccatgagttccagaagaagcatatgaccaga 1120
DB 2551 AGGACACAGAGAGGAGGAGGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2610
QY 1121 agtgcagagtcacatgatgtgtgacagggcccttcgagcgtgtgtgagttgccaatggccacc 1180
DB 2611 AGGGGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2670
QY 1181 agaaagcaagtagaagaagacccaagaacagcagtcgcaagagttcagacacagagag 1240
DB 2671 AGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2727
QY 1241 cagaaagagttctacagcnaaatatcgaaacaaatgagagagcagagccagagttggagagc 1300
DB 2728 AGGAGGGGACAGAGGAGGAGGAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2787
QY 1301 agagagcaccggaactaccgtgtgagggcttcaggttcaggaagtttgaacgggtccaccatcc 1360
DB 2788 AGGAGGGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2847
QY 1361 tccgcaatgccctgtgtgtgacatgtaaccagctctcactgagttgtgtcaagagatgagt 1420
DB 2848 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2907
QY 1421 agctctatgaggaagtcggtgtgaccttgagg 1453
DB 2908 AGGGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2940

RESULT 12
PCT-US93-04648-15
Sequence 15, Application PC/TUS9304648

GENERAL INFORMATION:
APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Natalie A., Mark, Melani
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04648
FILING DATE: 19930517
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/884811
FILING DATE: 18-MAY-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/885971
FILING DATE: 18-MAY-92
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 755,779P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
PCT-US93-04648-15

Query Match 3.0%; Score 63.4; DB 6; Length 10596;
Best Local Similarity 43.7%; Pred. No. 7.6e-08;
Matches 329; Conservative 0; Mismatches 421; Indels 3; Gaps 1;

QY 701 tccgagatgctcttggtgacggaacttcacgcccacacaggtatgagtgtactgc 760
DB 2191 TCCAGATTGCAATTGGCTGCCAAGGGACCCACGGTGGAAACAGGAGCAGAGCAGGAGCAGCGG 2250
QY 761 agaggtctctgagcagcaggaagatgtcaagatgtgaggaagctgtctcagacagagg 820
DB 2251 AGGGGACAGAGAGGAGGAGGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2310
QY 821 cccagcggaggaagaggtacggaagagctgtgtcagattgcaacgagctgtgtgc 880
DB 2311 AGGGGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2370
QY 881 agacagagatgaattcccttgaggaacctcttgaacctcctgaagcagaacagagaatg 940
DB 2371 AGGAGGGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2430
QY 941 tggagcagtcacacatccagctggccctggccctgctgtgaggaagctgtgagccctggagg 1000
DB 2431 GGCAGGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2490
QY 1001 agttccgagagagacagaagaagcagcggaaagatgtatgaggccatctgagccgtgtcc 1060
DB 2491 AGCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2550
QY 1061 agaagagcaagttgtcgtctctacaagaagaccatgagttccagaagaagcatatgaccaga 1120

Mon Sep 18 09:04:28 2000

us-09-068-377-2.ini

Page 12

Search completed: September 15, 2000, 10:03:44
Job time: 2879 sec

```
51  GluArgTyrGlyLysGluLeuValGlnIleAlaArgLysAlaGlyGly 67
|||||
832  GAGAGTACGGAGGAGCGTGTGCAGATTGCACCAAGCGTGGTGGCA 881
|||||
67  nThrGluMetAsnSerLeuArgThrSerPheAspSerLeuLysGlnIle 84
|||||
882  GACAGAGTGAATTCCTGAGGACCTCTTTGACTCCCTGAAGCAGCAAA 931
|||||
84  hrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaArgGlu 100
|||||
932  CAGAGAAATGTGGCAGTGCACACATCCAGCTGGCCCTGGCCCTGCTGAG 981
|||||
101  GluLeuArgSerLeuGluGluPheArgGluArgGlnLysGluArgIle 117
|||||
982  GAGCTGGCGAGCGCTGGAGGAGTTCGAGAGAGACAGAAAGAGCAGCGAA 1031
|||||
117  slsYtyrGluAlaIleMetAspArgValGlnLysSerLysLeuSerLeu 134
|||||
1032  GAAGTATGAGGCCATCATGGACCGTGTCCAGNAGAGCAAGTTGCGCTCT 1081
|||||
134  yLysLysThrMetGluSerLysLysAlaTyrAspGlnLysCysArgAsp 150
|||||
1082  ACAAGAAGACCATGAGTCCAAAGAGCATATGACCAGAAGTGCAGGAT 1131
|||||
151  AlaAspAspAlaGluGlnAlaPheGluArgValSerAlaAsnGlyHisG 167
|||||
1132  GCAGATGATGTGACAGCGCTTCAGGCGTGTGAGTGCCATGGCCACCA 1181
|||||
167  nLysGlnValGluLysSerGlnAsnLysAlaLysGlnCysLysGluSerA 184
|||||
1182  GAAGCAAGTAGAAGAGCCAGACAAAGCCAGCAGTGCAGAGAGTCAG 1231
|||||
184  laThrGluAlaGluArgValTyrArgGlnAsnIleGluGlnLeuGluArg 200
|||||
1232  CCACAGAGCGCAAGAGAGTGTACAGGCAAAATATCGAACTCGAGAGA 1281
|||||
201  AlaArgThrGluTyrGluGlnGluHisArgThrCysGluAlaPheG 217
|||||
1282  CGGAGGACCGAGTGGGAGCAGCAGCCGGACTACCTGTGAGGCGCTTCCA 1331
|||||
217  nLeuGlnGluPheAspArgLeuThrIleLeuArgAsnAlaLeuTyrValH 234
|||||
1332  GTTGCAGGAGTTTGACCGGCTCACCATCTCCGCAATGCCCTGTGGGTGC 1381
|||||
234  isCysAsnGlnLeuSerMetGlnCysValLysAspAspGluLeuTyrGlu 250
|||||
1382  ACTGTACCCAGCTCTCCATGCAGTGTGTCAAGGATGATGAGCTCTATGAG 1431
|||||
251  GluValArgLeuThrLeuGluGlyCysAspValGluGlyAspIleAsnG 267
|||||
1432  GAAGTGGCGCTGACCTTGAGGGCTGTGATGTGAAGTGACATCAATGG 1481
|||||
267  yPheIleGlnSerLysSerThrGlyArgGluProProAlaProValProt 284
|||||
1482  CTTTCATCCAGTCCAAAGAGCACTGCGCAGAGAGCCCCAGCTCCGGTGCCT 1531
|||||
284  yrGlnAsnTyrTyrAspArgGluValThrProLeuIleGlySerProSer 300
|||||
1532  ATCAGAACTACTATGACAGGAGGTGACCCCACTGATTGGCAGCCCTAGC 1581
|||||
301  ileGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisG 317
|||||
1582  ATCCAGCCCTCCCTGCGGTGTGATAAAGAGTTCTCTGGGCTGCTACATGG 1631
|||||
317  ySerProLysThrThrProSerAlaProAlaAlaSerThrGluThrLeu 334
|||||
1632  AAGTCCCAAGACCAACACTTCTCTCTCTGCTGCTTCCACAGAGACTCTGA 1681
|||||
334  hrProThrProGluArgAsnGluLeuValTyrAlaSerIleGluValGln 350
|||||
1682  CTCCCACCCCTGAGCGGAGTGTGTTGTTGTCTACGCATCCATCGAAGTGCAG 1731
|||||
```

```
351  AlaThrGlnGlyAsnLeuAsnSerSerAlaGlnAspTyrArgAlaLeuTy 367
|||||
1732  GCGAGCCAGGGAAACCTTAACCTCATCAGCCAGGACTACCGGGCACTCTA 1781
|||||
367  rAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspI 384
|||||
1782  CGACTACACTGCACAGAATTCTGATGAGCTGGACATTTCCGCGGAGACA 1831
|||||
384  leLeuAlaValIleLeuGluGlyGluAspGlyTyrTrpThrValGluArg 400
|||||
1832  TCTTGGCGGTCTATCTCGAAGGGAGGATGGCTGTGACTGTGGAGCGG 1881
|||||
401  AsnGlyGlnArgGlyPheValProGlySerTyrLeuGluLysLeu 415
|||||
1882  AACGACAAACGTGGCTTTGTCCCTGGGTCGTACTTGGAGAAAGCTC 1926
|||||
seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-938-830-28
seq_documentation_block:
; Sequence 28, Application US/08938830
; Patent No. 6040437
; GENERAL INFORMATION:
; APPLICANT: Lasky, Laurence A.
; APPLICANT: Dowdenko, Donald J.
; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
; TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPs)
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,830
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/798419
; FILING DATE: 07-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1066P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1613 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-938-830-28
```

```
alignment_scores:
  Quality: 1714.50      Length: 434
  Ratio: 4.684          Gaps: 3
  Percent Similarity: 84.332  Percent Identity: 77.880
```

```
alignment_block:
```

```
US-09-068-377-1 x US-08-938-830-28 ..
```

```
Align seg 1/1 to: US-08-938-830-28 from: 1 to: 1613
```

```
1 MetMetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheTh 17
```

```
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
344 ATGATGCCCGACCTCAGTTCAGATCAAGATGCTTTTGGTCAGGAGCTTCAC 393
17 rAlaHisThrGlyTyrGluValLeuLeuGlnArgLeuLeuAspGlyArgL 34
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
394 AGCCACACAGGGCTACGAGGTGCTGCTCAGCGGCTTCTGGATGGCAGA 443
34 ysMetCysLysaspValGluGluLeuLeuArgGlnArgAlaGlnAlaGlu 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
444 AGATGTGCAAGACATGGAGGAGCTACTGAGGACAGAGGGGCCAGCGGAG 493
51 GluArgTyrGlyLysGluValGlnIleAlaArgLysAlaGlyGlyCl 67
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
494 GAGCGGTACGGGAAGAGCTGGTGAGATCGACCGAAGGAGGCTGGCCA 543
67 nThrGluMetAsnSerLeuArgThrSerPheAspSerLeuLysGlnGln 84
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
544 GACGGAGATCAACTCCCTGAGGCGCTCCTTGACCTCTTGAAGACGACAA 593
84 hrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
594 TGGAGAATGTGGCAGCTCACACATCCAGCTGCCCTGACCCCTGCGGTGAG 643
101 GluLeuArgSerLeuGluPheArgGluArgGlnLysGluGlnArgLys 117
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
644 GAGCTGCGGAGTCTCGAGGAGTTCTGTGAGAGGCAGAGGACAGAGGAA 693
117 sLys ..... 118
|||||
694 GAAGGCGATGGCTGTCGCGAGACAGAGTCAGTCATGGAAGTGAAGTCCC 743
119 .....TyrGluAlaIleMetAspArgValGlnLysSerLysLeu 131
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
744 CATCATGGGAGTATGAGGCGCTCATGAGCGGGTCCAGAAGAGCAAGCTG 793
132 SerLeuTyrLysThrMetLysSerLysLysAlaTyrAspGlnLysCy 148
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
794 TCGCTCTACAAGAGCGCATGGAGTCCAAAGAACATACGACGACAAGTG 843
148 sArgAspAlaAspAlaGluGlnAlaPheGluArgValSerAlaAsnG 165
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
844 CCGGACCGGACGACGCGGAGCAGCGCTTCGAGCGCATACCGCCAAGC 893
165 lyHisGlnLysGlnValGluLysSerGlnAsnLysAlaLysGlnCysLys 181
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
894 GCCACCAAGCAGGTGGAGAAGAGTCAGACAAAGCCAGGCAAGTCAAG 943
182 GluSerAlaThrGluAlaGluArgValTyrArgGlnAsnIleGluGlnLe 198
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
944 GACTCGGCCCGGACGAGCGGGTATACAGGCAGAGCATTTGCCAGCT 993
198 uGluArgAlaArgThrGluTyrGluGlnGluHisArgThrThrCysGluA 215
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
994 GGAGAAGTCCGGCTGAGTGGGACAGGACGACCGGACCACTTGAGG 1043
215 laPheGlnLeuGlnGluPheAspArgLeuThrIleLeuArgAsnAlaLeu 231
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1044 CCTTTCAGCTGCAAGAGTTTGACCGGCTGACCATTTCCGCAACGCCCTG 1093
232 TrpValHisCysAsnGlnLeuSerMetGlnCysValLysAspAspGluLe 248
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1094 TGGGTGCACAGTAAACCACTCTCCATGAGTGTGTCAAGGATGATCAGCT 1143
248 uTyrGluGluValArgLeuThrLeuGluGlyCysAspValGluGlyAspI 265
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1144 CTACGAGGAGTGGCGCTGACGCTGGAGGCTGCAGCATACAGCCCGACA 1193
265 leAsnGlyPheIleGlnSerLysSerThrGlyArgGluProAlaPro 281
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1194 TCGACAGTTTCATCCAGCCCAAGAGCACGGGCACAGACCCCCC ..... 1237
282 ValProTyrGlnAsnTyrTyrAspArgGluValThrProLeuIleGlyse 298
```

```
1237 ..... 1237
298 rProSerIleGlnProSerCysGlyValIleLysArgPheSerGlyLeuL 315
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1238 .....AGTTTCTCTGGACTGC 1253
315 euHisGlySerProLysThrThrProSerAlaProAlaLaSerThrGlu 331
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1254 TGCACGGAAGTCCCAAGACCACTTCTGCTAGCTTCTGCTGGCTCCACAG 1303
332 ThrLeuThrProThrProGluArgAsnGluLeuValTyrAlaSerIleG 348
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1304 ACCCTGACCCCAACCCCGAGCGGAATAGGGTGTCTACACAGCATCCG 1353
348 uValGlnAlaThrGlnGlyAsnSerSerSerAlaGlnAspTyrArgA 365
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1354 AGTGACGAGGATACAGGGAACCCGCGCTCACCGCCAGGACTACCGGG 1403
365 laLeuTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSerAla 381
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1404 CGCTCTACGATTATACAGCGCAGAACCCAGATGAGCTGGACCTGTCCGCG 1453
382 GlyAspIleLeuAlaValIleLeuGluGlyGluAspGlyTrpTrpThrVa 398
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1454 GGAGAC .....ATCCTGGAAGGGGAGGATGGCTGGTGGACTGT 1491
398 lGluArgAsnGlyGlnArgGlyPheValProGlySerTyrLeuGluLysL 415
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1492 GGAGAGGAACGGCAGCGTGGCTTCGTCCTCGTCTCTACCTGGAGNAGC 1541
415 eu 415
||
1542 TT 1543
seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:us-08-435-454-3
seq_documentation_block:
; Sequence 3, Application US/08435454
; Patent No. 5605830
; GENERAL INFORMATION:
; APPLICANT: Prendergast, George C.
; APPLICANT: Sakamuro, Daitoku
; TITLE OF INVENTION: Murine and Human C-Myc Interacting
; TITLE OF INVENTION: Protein and Uses Therefor
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P O Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/435,454
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST6005A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1925 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

[illegible]

```

seq_documentation_block:
; Sequence 5, Application US/08919145
; Patent No. 5958753
; GENERAL INFORMATION:
; APPLICANT: Prendergast, George C.
; TITLE OF INVENTION: Bau, A Bin1 Interacting Protein, and
; TITLE OF INVENTION: Uses Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA

```

```

; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,145
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/025,482
; FILING DATE: 29-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kodroff, Cathy A.
; REGISTRATION NUMBER: 33,980
; REFERENCE/DOCKET NUMBER: WST73AUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1925 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 60..1412
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 60
; US-08-919-145-5

alignment_scores:
  Quality: 150.50      Length: 461
  Ratio: 0.660        Gaps: 20
  Percent Similarity: 49.458      Percent Identity: 21.692

alignment_block:
  US-09-068-377-1 x US-08-919-145-5      ..
  Align seg 1/1 to: US-08-919-145-5      from: 1 to: 1925

33  ArgLysMetCysLysAspValGluGluLeuLeuArgGlnArgAlaGlnAl 49
108 AAGAAGCTCACCCGCGCAGGAGAGGTTCTCCAGAAAGCTGGGAAGGC 157
49  aGluGluArgTyrGlyLysGluValGlnIleAlaArgLysAlaGlyG 66
158 AGATGAGACCAAGGATGAGCAGTTTGAGCAGTGCCTCCAGAAATTTCAACA 207
66  IyGln...ThrGlu.....MetAsnSerLeuArgThrSerPhe 77
208 AGCAGCTGCGGAGGCGCACCCGCTGCAGAGGATCTCGGACCTACCTG 257
78  AspSerLeuLysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLe 94
258 GCCTCCGTCAGAACGATGACGAG..... 281
94  uAlaLeuAlaLeuArgGluGluLeuArgSerLeuGluPheArgGluA 111
282 .....GCTTCCAGAAAGCTGAATGATGCTGCAGGAGGTTATGAGC 324
111 rgGlnLysGluGlnArgLysLysTyrGluAlaIle..... 122
325 CGGATTGGCCCGCAGGATGAGGCAACAGATCGCAGAGAAACAACGAC 374
123 .....MetAspArgValGlnLysSerLysLeuSerLeuTyrLysLy 136
375 CTGCTGTGATGATTACCACCAAG.....CTGGTGGAGCCA 412

```

```

136 sThrMetGluSerLysLysAlaTyr.....AspGlnLysC 148
413 GCGCTGCTGACCATGGACACGTAAGTTCCTGCGCAGTTCCCGCATCAAGT 462
148 ysArgAspAlaAspAlaGluGlnAlaPheGluArgValSerAlaAsn 164
463 CACGCTATTCGCAAGCGGGCGCAAGCTGGTGATACGACATGCGCCG 512
165 GlyHisGlnLysGlnValGlu.....LysSerGlnAsnLysAl 177
513 CACCTAGTAGTCTCCCTTCAAACTGCCAAAAGAGGATGAAGCCAAAAT 562
177 aLysGlnCysLysGluSerAlaThrGluAlaGluArgValTyrArgGln. 193
563 TGCCAAGCGCGGAGGAGCTCATCAAGCCAGAAAGGTGTTTGAGGAGA 612
194 ..AsnIleGluGlnLeuGluArgAlaArgThrGluTrpGluGlnGluHis 209
613 TGAATGTGGATCTGCAGGAGGAGCTGCTCCCTGTGGAACAGCCCGTA 662
210 ArgThrThrCysGluAlaPheGlnLeu..... 218
663 GGTTCCTAGCTCAACACGTTCCAGAGCATCGCGGGCTGGAGAAACTT 712
219 ...GlnGluPheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValH 234
713 CCACAAGGAGATGAGCAAGCTCAACACCAAGCTCAATGATGCTGTGTCG 762
234 isCysAsnGln.....LeuSerMetGlnCys 242
763 GCCTGGAGAAGCAACAGGAGCAACACCTTCACGCTCAAGGCCAGCC 812
243 ValLysAspGluLeuTyrGluGluValArgLeuThrLeuGluGlyC 259
813 AGAAAGAAAGTAACTGTTTCGCGCTGCGCAGAAAGAACAGACAGTGA 862
259 sAspValGluGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyA 276
863 CAACGCGCTGCAGAAAGGAAC.....AAGAGCCTTCGCTCCAG 903
276 rgGluProAlaProValProTyrGlnAsnTyrTyrAspArgGluVal 292
904 ATGGCTCCCTGCGCCACCC...GAGATCAGAGTCAACACGAGCCA 950
293 ThrProLeuIleGly.....SerProse 300
951 GAGCGCGCGCGGCGCACGCGCGGCGCACCTCCCAAGTCCCATC 1000
300 r.....IleGlnProSerCysG 306
1001 TCAGCCAGCAGAGGCTCGAGGTGGCGTGGGACCCCAACCTGCGGCTG 1050
306 IyValIleLysArgPheSerGlyLeuLeuHisGlySerProLysThr 322
1051 GAGCC.....CAGGAGCCAGCGGAGACTTCTCAAGTGAAGCAGCTCC 1094
323 ProSerAlaProAlaSerThrGluThrLeuThrProThrProGluAr 339
1095 ACCTCTCTCTCTGCTGCTGTGTGGAGAGCTTCCCGAGCAACT..... 1136
339 gAsnGluLeuValTyrAlaSerIleGlnValGlnAlaThrGlnGlyAsnL 356
1137 .....GTGAATGGCACCGTGGAGGCGCAGTGGGCGCGCGCT 1176
356 euAsn.....SerSerAlaGlnAspTyrArgAlaLeuTyrAspTyr 369
1177 TGGACCTGCCCGCAGGTTTCATGTTCAAGGTACAGGCCCGCAGCAGCTAC 1226
370 ThrAlaGlnAsnSerGluLeuAspIleSerAlaGlyAspIleLeuAl 386
1227 ACGGCACCTGACACAGACGAGCTGCATGCTCAAGGCTGGTGTGTGCT 1276

```

```

386 avallle.....LeuGluGlyGluAspGlyTrp..... 395
||||| ||||| ||||| ||||| |||||
1277 GGTGATCCCTTCCAGAACCCCTGAAGACAGGATCAAGGCTGGCTCATGG 1326

396 .....TrpThrValGluArgAsnGlyGln.....Arg 404
||||| :|||: :|||: :|||: :|||:
1327 GCGTGAAGGAGGCGCACTGGACACCAGCACAAGAAGCTGGAGAGTGCCGT 1376

405 GlyPheValProGlySerTyrLeuGluLysLeu 415
||| :|||: ||||: ||||: ||||:
1377 GGCGTCCTCCCCGAGAAGCTTCACGTGAGAGGGTC 1409

```

seq_name: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US96-06231A-3

```

seq_documentation_block:
; Sequence 3, Application PC/TUS9606231A
; GENERAL INFORMATION:
; APPLICANT: Wistar Institute of Anatomy & Biology
; TITLE OF INVENTION: Murine and Human Box-Dependent
; TITLE OF INVENTION: Myc-Interacting Protein (BIN1) and Uses Therefor
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P O Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/06231A

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/435,454
FILING DATE: 05-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST60A0PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1925 base pairs

```

/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: unknown
/ MOLECULE TYPE: CDNA
/ FEATURE:
/   NAME/KEY: CDS
/   LOCATION: 60..1412
/ FEATURE:
/   NAME/KEY: mat_peptide
/   LOCATION: 60
PCT-US96-06231A-3

```

```
alignment_scores:
  Quality: 150.50
  Ratio: 0.660
  Percent Similarity: 49.458
  Gaps: 20
  Percent Identity: 21.692
  Length: 461
```

```

alignment_block:
  US-09-068-377-1 x PCT-US96-06231A-3
  ..
  Align seg 1/1 to: PCT-US96-06231A-3 from: 1 to: 1925

```

33 ArgLysMetCysLysAspValcIuGluLeuArgGlnArgAlaGlnAl 49
108 AAGAAGCTTCAACCGCGCCAGGAAAGTTCTCCACAGCTGGGAAGC 157
49 aGluGluArgTyrGlyLysGluLeuValcIuGlnIleAlaArgLysAlaGlyG 66
158 AGATGACCAACGATGACAGTGTGACGATGGCGTCCAGAATTTCAACA 207
66 lyGln...ThrGlu.....MetAsnSerLeuArgThrSerPhe 77
208 AGCAGCTGACGAGGACCCCGGCTGCAGAGGATCTCCGACCTACCTG 257
78 AspSerLeuLysGlnGlnThrGluAsnValGlySerAlaHisIleGln 94
258 GCCTCCGTCAAAGCCATCAGAG..... 281
94 uAlaLeuAlaLeuArgGluGluLeuArgSerLeuGluGluPheArgGlu 111
282GCTTCCAAAGAGCTGAATGAGTGTCTGCAGGAGGTGTATGAGC 324
111 rgGlnLysGluGlnArgLysLysTyrGluAlaIle..... 122
325 CCGATTGCCCGGCGGATGAGGCAACAAGATCCAGAGCAACAGCAG 374
123MetAspArgValcIuLysSerLysLeuSerLeuTyrLysLy 136
375 CTGCTGTGGATGGATTACCACCAAG.....CTGCTGGACCA 412
136 sThrMetCysLysLysAlaTyr.....AspGlnLysC 148
413 GGCCTGCTGACCATGACAGCTACCTGGCGCAGTTCGCCGACATCAAGT 462
148 ysArgAspAlaAspAlaGluGlnAlaPheGluArgValSerAlaAsn 164
463 CACCGATTGCCAACCGGGGGCCAAAGCTGTGGACTACGACAGTCCCGGG 512
165 GlyHisGlnLysGlnValGlu.....LysSerGlnAsnLysAl 177
513 CACCACTACGAGTCCCTTCAAACTGCCAAAAGAGGATGAAGCCAAAAT 562
177 aLysGlnCysLysGluSerAlaThrGluAlaGluArgValTyrArgGln. 193
563 TGCCAAAGCGGAGGAGAGCTCATCAAGCCCAAGAGCTGTGTGAGGAGA 612
194 ..AsnIleGluGlnLeuGluArgAlaArgThrGluTrpGluGlnGluHis 209
613 TGAATGTGGATCTCCAGGAGGAGTGGCTCCCTGTGGAAACAGCGCGCTA 662
210 ArgThrThrCysGluAlaPheGlnLeu..... 218
663 GGTWTCTACGTCAACACGTTCCACAGCATCGCGGCTGGAGGAAAACCT 712
219 ...GlnGluPheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValH 234
713 CCACAAGGAGATGACGAAGCTCAACACAGACCTCAATGATGTGTGTGCTG 762
234 iCysAsnGln.....LeuSerMetGlnCys 242
763 GCCTGGAGAGCAACACCGGAGCAACACCTTCACGGTCAAGCGCCAGCCC 812
243 ValLysAspGluLeuTyrGluGluValArgLeuThrLeuGluGlyCys 259
813 AGAAGAAAAAGTAAACGTTTTCCGCGCTGGCAGCAAAAGAACACAGTGA 862
259 aspValGluGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyA 276
863 CAACGCGCTTGCAAAAGGGAAAC.....AAGAGCCCTTCGCGCTCCAG 903
276 rgGluProProAlaProValProTyrGlnAsnTyrTyrAspArgGluVal 292
904 ATGGCTCCCTTCGCCGCCACCCC...GAGATCAGAGTCAACCAAGCGCCA 950
293 ThrProIleGly.....SerProse 300


```
282 .....GCTTCCAAGAGCTGAATGAGTCTCTGCAGGAGGTGTATGAGC 324
111 rglnLysGluInArgLysLysTyrGluAlaIle..... 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
325 CCGATTGGCCCGCGGATGAGGCAAAACAAGATCGCAGAGACAAGCAC 374
123 .....MetAspArgValGlnLysSerLysLeuSerLeuTyrLys 136
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
375 CTGCTGTGGATGGATTACCAACCAAG.....CTGGTGGACCA 412
136 sThrMetGluSerLysLysAlaTyr.....AspGlnLysC 148
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
413 GCGCTGCTGACCATGACACGTACCTGGGCCAGTTCCTCCCGACATCAAGT 462
148 ysArgspAlaAspAlaGluGlnAlaPheGluArgValSerAlaAsn 164
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
463 CAGGCATTGCAAGCGGGCGCAAGCTGTGGTACGACAGTGCCTGG 512
165 GlyHisGlnLysGlnValGlu.....LysSerGlnAsnLysAl 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
513 CACCATACGAGTCCCTTCAAACTGCCAAAAGAGATGAAGCCAAAT 562
177 aLysGlnCysLysGluSerAlaThrGluAlaGluArgValTyrArgGln 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
563 TGCCAAAGCGCGAGGAGTCTCATCAAGCCCAAGAGGTGTGTGAGGAGA 612
194 ..AsnIleGluInLeuGluArgAlaArgThrGluTrpGluGlnHis 209
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
613 TGAATGTGGATCTGCAGAGGAGCTGCCCTCCTGTGGACACGCCGGTA 662
210 ArgThrThrCysGluAlaPheGlnLeu..... 218
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
663 GGTTTCTACGTCAACAGTTCAGAGCATCGCGGGCTTGAGGAAACTT 712
219 ....GlnGluPheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValH 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
713 CCACAGGAGATGAGCAAGCTCAACAGAACCTCAATGATGTGCTGTG 762
234 isCysAsnGln.....LeuSerMetGlnCys 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
763 GCCTGGAGAGCAACACGGGAGCAACACCTTCACGGTCAAGGCCACGCC 812
243 ValLysAspGluLeuTyrGluGluValArgLeuThrLeuGluGlyC 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
813 AGAAAGAAAGTAACTGTTTCGCGGCTGCGCAGAAAGAACACAGTGA 862
259 sAspValGluGlyAspIleAsnGlyPheIleGlnSerLysSerThrGly 276
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
863 CAACGGCCTGCCAAAAGGGAAC.....AAGAGCCCTTCGCCCTCCAG 903
276 rgGluProAlaProValProTyrGlnAsnTyrTyrAspArgGluVal 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
904 ATGGCTCCCTGCGGCCACCC...GAGATCAGAGTCAACACAGGACCA 950
293 ThrProLeuIleGly.....SerProSe 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
951 GAGCGCGCGCGGGGCCACGCGCGGGGCCACCTCCCAAGTCCGCCATC 1000
300 r.....IleGlnProSerCysC 306
1001 TCAGCCAGCAGAGGCTCGGAGTGGCGGGTGGACCAACCTCGCGGTG 1050
306 lyValIleLysArgPheSerGlyLeuHisGlySerProLysThrThr 322
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1051 GAGCC.....CAGGAGCCAGGAGGAGCGCGCAAGTGAAGCAGCGCTCC 1094
323 ProSerAlaProAlaSerThrGluThrLeuThrProThrProGluAr 339
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1095 AGCTCTCTCTCTGCTGCTGTGGTGGAGACCTTCCACAGCAACT..... 1136
339 gAsnGluLeuValTyrAlaSerIleGluValGlnAlaThrGlnGlyAsnL 356
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1137 .....GTGAATGGCACCCTGGAGCGGCGCAGTGGGGCGGGCGCT 1176
```

```
356 euAsn.....SerSerAlaGlnAspTyrArgAlaLeuTyrAspTyr 369
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1177 TGGACCTGCCCCAGGTTTCATGTTCAAGGTACAGCCAGCCAGCAGTAC 1226
370 ThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspIleLeuAl 386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1227 ACGCCCACTGACAGACGAGCTGACGCTCAAGCTGCTGATGTGTGCT 1276
386 aValIle.....LeuGluGlyGluAspGlyTrp..... 395
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1277 GGTGATCCCTTCCAGAACCTGAAGACGAGATGAAGGCTGGCTCATGG 1326
396 .....TrpThrValGluArgAsnGlyGln.....Arg 404
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1327 GCGTGAAGGAGGAGCGACTGGAACACGACACAGAAGCTGGAGAAAGTGC 1376
405 GlyPheValProGlySerTyrLeuGluLysLeu 415
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1377 GCGCTCTTCCCGAGAACTTCACTGAGAGGGTC 1409
seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-878-563A-2
seq_documentation_block:
; Sequence 2, Application US/08878563A
; Patent No. 5891674
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: INSULIN RECEPTOR TYROSINE KINASE SUBSTRATE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,563A
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0323 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2080 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRSTNOT04
; CLONE: 918158
; US-08-878-563A-2
```

```
alignment_scores:
  Quality: 148.50      Length: 439
  Ratio: 0.678        Gaps: 21
  Percent Similarity: 49.886  Percent Identity: 21.640
```

alignment_block:

```
US-09-068-377-1 x US-08-878-563A-2
Align seg 1/1 to: US-08-878-563A-2 from: 1 to: 2080

35 MetCysLysAspValGluGluLeuLeuArgGlnAlaGlnAlaGluGlu 51
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
195 ATGGGAAGAAATTACGAGAGAGGACTGCGAGGTGTGAGTATGAGCAA 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 uArgTyrGlyLysGluLeuVal.....GlnIleAlaArgLysAlaG 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
245 AGGTACTTTGACCCCTGGTGAAGATGGGGAGCTGGCCAGCAGAGCC 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 lyGlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeuLys 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 AGGCTCAAAAGAACTCGGAGACGTTCTTCCAGATGGCTGAAGTCCAC 344
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
82 GlnGlnThrGluAsn.....ValGlySerAlaHisIleGlu 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345 AGGCAGATCCAGAAATCAGCTGGAAGAATGCTGAAGTCTTTTCACAAGA 394
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
93 nLeuAlaLeuAlaLeuArgGluGluLeuArgSerLeuGluGluPhe.... 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
395 GCTGCTACGAGCTGGAGCAGAGGAGGAGCTGGACTCCAGGTATCTGA 444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
109 .....ArgGluArgGlnLysGluGlnArgLysLysTyrGluAla 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
445 GTGCTGCGCTAAAGAAATACAGACTGAGCAAGAGGAGCAAGGCGAGCC 494
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
122 IleMetAspArgValGlnLysSerLysLeuSerLeuTyrLysThrMe 138
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
495 ...CTGGACAAGTGTACGCTGAGCTGAAGAAGCTTGGGAAGAAGACCA 541
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
138 tGluSerLys.....LysAlaTyrAspGlnLysCysArgAspAlaA 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
542 GGGCAGCAAGAATCCTCAGAAAGTACTCGGACAGGAGTGCAGTACATCG 591
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
152 sp.....AspAlaGluGlnAlaPheGluArgValSerAlaAsnGly 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
592 AGCCATCAGCAACAAGCAGGCGAGCTGGAGAATTACGTGTCGCGAGGC 641
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
166 HisGlnLysGlnValGluLysSerGlnAsnLys.....Al 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
642 TACAAGACCGCACTGACAGAGGAGTGCAGGCGTCTCTCTCTCTCTCT 691
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
177 aLysGlnCysLysGluSerAlaThrGluAlaGluArgValTyrArgGlnA 194
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
692 GAACGAGTGCCTGCGCAAGAGTGCCTGTGGCAACAG..... 779
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
194 snIleGluGlnLeuGluArgAlaArgThrGluTrpGluGlnGluHisArg 210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
739 GCAAGGAGCTGCTCCGCAAGAGTGCCTGTGGCAACAG..... 779
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
211 ThrThrCysGluAlaPheGlnLeuGlnLysPheAspArgLeuThrIleLe 227
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
780 ...GCCTGTGCC.....GACCCCAAGCAAGATCCC 805
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
227 uArgAsnAlaLeuTrpValHisCysAsnGlnLeuSerMetGlnCysValL 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
806 GGAGCGCGCGGTG.....CAGCTCATGCAAGAGTGGCCCA 840
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
244 ysAspAspGluLeuTyrGluGluValArgLeuThrLeuGluGlyCysAsp 260
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
841 GCAACGGC.....GCCACCTCCCGAGCGCCCTG 869
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
261 ValGluGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyArgGlu 277
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
870 TCGGCCCTCCCAAGTCCAACTGCTGATTTCCGACCCCATTCGGGGGCCAA 919
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
277 uProProAlaProValProTyrGlnAsnTyrTyrAspArgGluValThrP 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
920 G...CCCTGCGCGGTGCC.....CCCGAGCTGGCAC 948
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

seq_name: /cqn2_5/ptodata/1/lna/5A_COMB.seq:US-08-466-390-3

seq_documentation_block:

```
; Sequence 3, Application US/08466390
; Patent No. 5686562
; GENERAL INFORMATION:
; APPLICANT: TOKUATLY, GARY
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,390
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ. EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
```



```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6306
; PUBLICATION INFORMATION:
; AUTHORS: COMPTON, DUANE A
; AUTHORS: SZILAK, ILYA
; AUTHORS: CLEVELAND, DON W
; TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR
; TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
; TITLE: SEGREGATION OF PROTEINS AT MITOSIS
; JOURNAL: J. Cell Biol.
; VOLUME: 116
; PAGES: 1395-1408
; DATE: 1992
US-08-467-781-3

alignment_scores:
    Quality: 147.50          Length: 457
    Ratio: 0.706             Gaps: 22
Percent Similarity: 45.733   Percent Identity: 22.101

alignment_block:
US-09-068-377-1 x US-08-467-781-3 ..

Align seg 1/1 to: US-08-467-781-3 from: 1 to: 6306

19  HisThrGlyTyTyrGluValLeuLeuGlnArgLeuLeuAspGlyArgLysMe 35
    |||||:::
4048  CACACCAGCACACAGGCCCTGGTGAGTGAGCTGCTGCCAGCTAAGCACCT 4097
    :::::

35  tCysLysaspValGluGluLeuLeuArgGlnArgAlaGlnAlaGluGluA 52
    :|||
4098  CTGC.....CAGCAGCTGCAGGCCAGCAGGCCGCTGCCGGAGAAAC 4138
    :::::

52  tqTyrGlyLysGluLeuValGlnIleAlaArgLysAlaGlyGly..... 66
    ||:::  ::|||::|  ||  :::  |||||::|

4139  GCCACCGTGAGAGCTGGAGCAGACGACGAGCCGCTGGGGACATGCCG 4188
    |||||  ::|||::|  ||  :::  |||||::|

67  .....GlnThrGluMet..... 70
    |||  |||:::

4189  GCAGAGCTGCTCGCGGCCCGCAGCGGGAGCTTGGGAGCTGATTCCTCTGCG 4238
    ::|||

71  .....AsnSerLeuArgThrSerP 77
    :::::|||||:::

4239  GCAGAAGGTGGCAGACGAGGAGCGAACAAGCTCAGCAGCTCGCGGCAGAGA 4288
    :::::

77  heAspSerLeuLysGlnGlnThrGluAsnValClySerAlaHisIleGln 93
    ||||  ::|||::|  :::  ::|||::|

4289  AGGCCAGCTATCGCAGAGCAGCTGAGCATGCTGAAGAAGCGCATGCCCTG 4338
    ||||  ::|||::|  :::  ::|||::|

94  LeuAlaLeuAlaLeuArgGluGlnLeuArgSerLeuGluGlu..... 107
    |||||  |||||  |||||::|  |||

4339  CTGGCA.....GAGGAGAACCGGGGGCTGGGTGAGCGGGCCAA 4376
    :::::

108  .....PheArgGluArgGlnLysGluGln...ArgLysLysT 119
    |||  |||  :::  ::|||  |||:::|

4377  CCTTGGCGCGCAGTTTCTGGAAGTGAGGTGGACCAAGCGCCGGGAAAAAGT 4426
    ::|||:::

119  yr.....GluAlaIleMetasp 124
    ||  |||:::  :::

4427  ATGTCCAAAGATTGGCAGCGGTACGTGTGATGCTGAGACCCGCTCTGGCT 4476
    ::|||:::

125  ArgValGlnLysSerLysLeuSerLeuTyrLys..... 135
    ::|||:::  |||  :::

4477  GAGGTGCAGCAGACGACACAGACACTGCCCGGAGCTGGAGTGATGAC 4526
    ::|||:::

136  .....LysThrMetGluSerLysLysAlaT 144
    |||:::  ::|||:::  :::

```

```
4527 TGCAAGTATGAGGTGCCAAGGTCAAGTCTCTGGAGGAGGCGAGCGGT 4576
144 yraspgln..... 146
4577 TCCAGGAGAGAGAGGACCAAACTACTGCCAGGTGGAGAACTCAGTAAG 4626
147 lycysargaspAlaaspAla.....GluGlnAlaPheG1 159
4627 AAAGTGGCTGACTGTGACCAAGCCAGCAAGGTGCAGCAGCAGCAAGCTGAA 4676
159 uArgValSerAlaasnGly.....HisGlnLysGlnValGluLysSerG 174
4677 GGCTGTCAGGCTCAGGAGCGGAGAGCCAGCAGGAGGCCGAGCGCTTCC 4726
174 InAsnLysAlaLysGlnCysLysGluSerAlaThrGluAlaGluArgVal 190
4727 AGCCCCAGCTGAATGAAGTCAACCCAGTTGAGCCAGAGGAGCAGGCA 4776
191 TyrArgGlnAsnIleGluGlnLeuGluArgAlaArgThrGluTrpGlu.. 206
4777 GCTGAGCACTATAAGCTGCAGATGAGAGCAAGCAACCAACATATTATGATGC 4826
207 .....GlnGluHisArgThrThrCysGluA 215
4827 CAAGAAGCAGCAGAACCAAGAGCTGCAGAGCAGCTGCGGAGCTGGAGC 4876
215 lapheGlnLeuGln.....GluPheAspArgLeuThr 225
4877 AGTGCACAGAAACAAACAGCTGCGAGCTGAAGCTGAACGCTGGGC 4926
226 IleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeuSerMetGlnCy 242
4927 .....CATGAGCTACAGCAGGCTGG 4946
242 sValLysAspAspGluLeuTyrGluGluValArg...LeuThrLeuGluG 258
4947 GCTGAAGACCAAGAGGCGTGAACAGCTGCCGCCACCTTACTGCCCAGG 4996
258 lYcysAspValGluGlyAspIle...AsnGlyPheIleGlnSerLysSer 273
4997 TGGCAGGCTGGAGCAGAGTGGCCATGTCAGACACGACGCTTCGAGAC 5046
274 ThrGly.....ArgGluProPr 279
5047 CTGGCAAAATTCAGGTGGCAACTGATGCTTTAAAGAGCCGTGAGCCCA 5096
279 oAlaProValProTyrGlnAsnTyrTyAsp.....A 290
5097 GGCTAAGCCCGCAGCTGGACTTGAGTATTGACAGCCTGGATGAGTCCG 5146
290 rgGluValThrProLeu...IleGlySer.....ProSerIleGlnPro 303
5147 AGGAGGGACCCCACTCAGTATCACCAGCAAGCTGCCTCGTACCAGCA 5196
304 SerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLy 320
5197 .....GACGGCAC 5204
320 sThrThrProSerAlaProAlaAlaSer...ThrGluThrLeuThrProT 336
5205 CAGGTCTCCCTGGAGNACAGGCTCACCATTATCTCCAGGCGCTGCCCCCA 5254
336 hrProGluArgAsnGluLeuValTyrAlaSerIleGluValGlnAlaThr 352
5255 AGGTAGATCCCTGGAGAGTCTCTACTTCACTCCATCCCTGCTCGGAGT 5304
353 GlnGlyAsnLeuAsnSerSer 359
5305 CAGGCCCTCCCTGGAGAGCAGC 5325
seq_name: /cgn2_6/ptodata/1/ina/5b_comb.seq:us-08-195-487-3
seq_documentation_block:
```

```
Sequence 3, Application US/08195487
Patent No. 5783403
GENERAL INFORMATION:
APPLICANT: TOUKATLY, GARY
APPLICANT: LIDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA HURWITZ & THIBEAULT
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/195,487
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/901,701
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESO, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/248-7000
TELEFAX: 617/248-7100
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6306 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..6306
PUBLICATION INFORMATION:
AUTHORS: COMPTON, DUANE A
AUTHORS: SZILAK, ILLYA
AUTHORS: CLEVELAND, DON W
TITLE: PRIMARY STRUCTURE OF OF NUMA, AN INTRANUCLEAR
TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
TITLE: SEGREGATION OF PROTEINS AT MITOSIS
JOURNAL: JOURNAL OF CELL BIOLOGY
VOLUME: 116
PAGES: 1395-1408
DATE: MAR-1992
US-08-195-487-3

alignment_scores:
Quality: 147.50 Length: 457
Ratio: 0.706 Gaps: 22
Percent Similarity: 45.733 Percent Identity: 22.101

alignment_block:
US-09-068-377-1 x US-08-195-487-3 ..
Align seg 1/1 to: US-08-195-487-3 from: 1 to: 6306
19 HisThrGlyTyrGluValLeuLeuGlnArgLeuLeuAspGlyArgLysMe 35
|||||: : : : : : : : : : : : : : : : : : : : : : : : : : :
4048 CACACCAACACACAGGCCCTGGTGTAGTGTGCTGCAGCTAAGCACCT 4097
35 tCysLysAspValGluGluLeuLeuArgGlnArgAlaGlnAlaGluGluA 52
```

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-483-924-3

```
;
;
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
;
; NAME/KEY: CDS
; LOCATION: 1..6306
; PUBLICATION INFORMATION:
; AUTHORS: COMPTON, DUANE A
; AUTHORS: SZILAK, ILYA
; AUTHORS: CLEVELAND, DON W
; TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR
; TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
; TITLE: SEGREGATION OF PROTEINS AT MITOSIS
; JOURNAL: J. Cell Biol.
; VOLUME: 116
; PAGES: 1395-1408
; DATE: 1992
; US-08-483-924-3
;
; alignment_scores:
;   Quality: 147.50      Length: 457
;   Ratio: 0.706        Gaps: 22
; Percent Similarity: 45.733 Percent Identity: 22.101
;
; alignment_block:
; US-09-068-377-1 x US-08-483-924-3 ..
;
; Align seg 1/1 to: US-08-483-924-3 from: 1 to: 6306
;
; 19 HistThrGlyTyrGluValLeuLeuGlnArgLeuLeuAspGlyArgLysMe 35
; 4048 CACACCAGCACAGGCCCTGGTGGTGGAGTGGCTGCCAGCTAAGCACCT 4097
;
; 35 tCysLysAspValGluGluLeuArgGlnArgAlaGlnAlaGluGluA 52
; 4098 CTGC.....CAGCAGCTGACAGCCGAGCAGCCGCTGCCGAGAAC 4138
;
; 52 rgTyrGlyLysGluLeuValGlnIleAlaArgLysAlaGlyGly..... 66
; 4139 GCCACCGTGAGGAGCTGGAGCAGCAGCAGCAGCAGCCGCTGGGGACTGCG 4188
;
; 67 .....GlnThrGluMet..... 70
; 4189 GCAGAGCTGCTGCGGCCCCAGCGGAGCTTGGGAGCTGATTCCTCTGCG 4238
;
; 71 .....AsnSerLeuArgThrSerP 77
; 4239 GCAGAGGTGGCAGCAGCAGGAGCGAAGCAGCTCAGCAGCTCGGGCAGAGA 4288
;
; 77 heAspSerLeuLysGlnThrGluAsnValGlySerAlaHisIleGln 93
; 4289 AGGCCAGCTATGCAGAGCAGCTGAGCATGCTGAAGAAGCGCATGGCCTG 4338
;
; 94 LeuAlaLeuAlaLeuArgGluGluLeuArgSerLeuGlu..... 107
; 4339 CTGGCA.....GAGGAGAACCGGGGGCTGGGTGAGCGGGGCCAA 4376
;
; 108 .....PheArgGluArgGlnLysGluGln...ArgLysLysT 119
; 4377 CCITGGCCGGCAGTTCTTGAAGTGGAGTTGGACCAGGCCCGGGAAAGT 4426
;
; 119 yr.....GluAlaIleMetAsp 124
; 4427 ATGTCCAAAGATTGGCAGCGCTAGTCTGATGCTGAGACCCGCTCTGGCT 4476
;
; 125 ArgValGlnLysSerLysLeuSerLeuTyrLys..... 135
; 4477 GAGTGCAGCAGAGAAGCAGACAGCAGCTGCCCCGGAGCTGGAGGTGATGAC 4526
;
; 136 .....LysThrMetGluSerLysLysAlaT 144
; 4527 TGCCAAGTATGAGGTTGCCAAGTCAAGGTCTTGGAGGAGCAGCGGCT 4576
```

```
144 yrAspGln..... 146
;
; 4577 TCCAGGAAGAGAGGCAGAACTCACTGCCAGGTGGAAGAACTCAGTAAG 4626
;
; 147 LysCysArgAspAlaAspAla.....GluGlnAlaPheG1 159
; 4627 AAAGTGGCTGACTCTGACCAAGCAGCAAGGTGCAGCAGCAGCAAGCTGAA 4676
;
; 159 uArgValSerAlaAsnGly.....HisGlnLysGlnValGluLysSerG 174
; 4677 GGCTGCCAGCTCAGGAGCGAGAGCCAGCAGGAGGCCAGCGCTTCC 4726
;
; 174 LnsnLysAlaLysGlnCysLysGluSerAlaThrGluAlaGluArgVal 190
; 4727 AGGCCAGCTGAATGAAGTGCAGCCAGTTGAGCCAGAGGAGCAGGCA 4776
;
; 191 TyrArgGlnAsnIleGluGlnLeuGluArgAlaArgThrGluTrpGlu... 206
; 4777 GCTGAGCATAAAGCTGCAGATGGAGAAAGCCAAACACATTATGATGC 4826
;
; 207 .....GlnGluHisArgThrThrCysGluA 215
; 4827 CAGAAGCAGCAGAACCAAGAGCTGCAGAGCAGCTCGGAGCGCTGGAGC 4876
;
; 215 laPheGlnLeuGln.....GluPheAspArgLeuThr 225
; 4877 AGCTGCAGAGGAAACAAAGAGCTGCAGCTGAAGCTGAACGCTGGC 4926
;
; 226 IleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeuSerMetGlnCy 242
; 4927 .....CATGAGCTACAGCAGGCTGG 4946
;
; 242 sValLysAspGluLeuTyrGluGluValArg...LeuThrLeuGluG 258
; 4947 GCTCAAGACCAAGAGGCTGAACAGAGCTGCCGCCACCTTACTGCCCG 4996
;
; 258 lyCysAspValGluGlyAspIle...AsnGlyPheIleGlnSerLysSer 273
; 4997 TGGCGACGCTGGAGGCACAGGTTGCCCATGCAGACAGCAGCTTCGAGAC 5046
;
; 274 ThrGly.....ArgGluProPr 279
; 5047 CTGGGCAAAATCCAGGTGGCAACTGATGCTTTAAAGAGCCGTGAGCCCA 5096
;
; 279 oAlaProValProTyrGlnAsnTyrTyrAsp.....A 290
; 5097 GGCTAAGCCCCAGCTGGACTTGAGTATTGACAGCGCTGGATCTGAGTGGC 5146
;
; 290 rgGluValThrProLeu...IleGlySer.....ProSerIleGlnPro 303
; 5147 AGGAGGGGACCCCACTCAGTATCACCAGCAAGCTGCTCGTACCCAGCCA 5196
;
; 304 SerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLy 320
; 5197 .....GACGGCAC 5204
;
; 320 sThrThrProSerAlaProAlaAlaser...ThrGluThrLeuThrProT 336
; 5205 CAGCGTCCCTGGAGAACCCAGCTCACCTATCTCCAGCGCTGCCGCCCA 5254
;
; 336 hrProGluArgAsnGluLeuValTyrAlaSerIleGluValGlnAlaThr 352
; 5255 AGGTAGAATCCCTGGAGAGTCTTACTTCACTCCCATCTCTGCTCGGAGT 5304
;
; 353 GlnGlyAsnLeuAsnSerSer 359
; 5305 CAGGCCCTCGGAGAGCAGC 5325
;
; seq_name: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-06160-3
; seq_documentation_block:
; ; Sequence 3, Application PC/TUS9306160
; ; GENERAL INFORMATION:
```



```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

5047 CTGGCAATTCAGGTGGCACTGATGCTTTAAAGAGCGGTGAGCCCA 5096
      |||
      |||
279 oAlaProValProTyrGlnAsnTyrTyrAsp.....A 290
      |||
      |||
5097 GGCTAAGCCCGCAGCTGGAGCTGAGTATGACAGCGCTGAGCTGCG 5146
      |||
      |||
290 rGluValThrProLeu...ileGlySer.....ProSerileGlnPro 303
      |||
      |||
5147 AGGAGGAGCCCGCAGTATGATACACAGCTGCTGACCGCCCA 5196
      |||
      |||
304 serCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLy 320
      |||
      |||
5197 .....GACGGCAC 5204
      |||
      |||
320 sThrThrProSerAlaProAlaAlaSer...ThrGluThrLeuThrProT 336
      |||
      |||
5205 CAGGCTCCCTGGAGAACGAGCTCACCTATCTCCAGCGCTGCCCGCCA 5254
      |||
      |||
336 hrProGluArgAsnGluLeuValTyrAlaSerileGluValGlnAlaThr 352
      |||
      |||
5255 AGGTAGATCCCTGGAGCTCTACTTCACTCCCATCCCTGCTCGGAGT 5304
      |||
      |||
353 GlnGlyAsnLeuAsnSerSer 359
      |||
      |||
5305 CAGGCCCGCTGGAGAGCAGC 5325
      |||
      |||
seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:us-08-691-814B-3
```

```

seq_documentation_block:
; Sequence 3, Application US/08691814B
; Patent No. 5981218
; GENERAL INFORMATION:
; APPLICANT: Rio, Marie-Christine
; APPLICANT: Tomasetto, Catherine
; APPLICANT: Basset, Paul
; APPLICANT: Byrne, Jennifer
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful
; as Leukemia Markers and in Breast Cancer Prognosis
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave, NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,814B
; FILING DATE: 31-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/002,183
; FILING DATE: 09-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1383.0090001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2543
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3846 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: 76..858
; US-08-691-814B-3

alignment_scores:
      Quality: 147.00      Length: 390
      Ratio: 0.821      Gaps: 13
      Percent Similarity: 45.897      Percent Identity: 20.000

alignment_block:
US-09-068-377-1 x US-08-691-814B-3 ..

Align seg 1/1 to: US-08-691-814B-3 from: 1 to: 3846

37 LysAspValGluGluLeuLeuArgGlnArgAlaGlnAlaGluArgTy 53
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27 CAGGACGCGGTGAGCTCAGCGCTCCCGCGCGCAAGATCGTGATCCCA 76
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
53 rGlyLysGluLeu.....ValGlnIleAlaArgLysAlaGlyGlnT 68
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77 TGAACCCCACTGCGCCCGGTGCGCGCAAGATCGTGATCCCA 126
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68 hrGluMetAsnSerLeuArgThrSerPheAspSerLeuLysGlnThr 84
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
127 GTGAACCTGCTGGATAAGTTCTGGCATAAAGCATGCTT..... 164
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85 GluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 101
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
165 .....CC 166
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 uLeuArgSerLeuGluGluPheArgGluArgGlnLysGluGlnArgLys 118
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 ATTCGAGACCTGCAAGATGACACTGAACATGAAGAACTACAAGGCTAC 216
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
118 ysTyrGluAlaIleMetAspArg...Val.GlnLysSerLysLeuSerLe 133
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217 GAGAAGAGCCCTACTGCACGACACACTTACCCCAAGCAGTCTTCCACCAT 266
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 uTyrLysLysThrMetGluSerLysLysAlaTyrAspGlnLysCysArg 150
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
267 GGTGGCGACACCCCGGAACCTTCGC..... 294
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150 spAlaAspAlaGluGlnAlaPheGluArgValSerAlaAsnGlyHis 166
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
295 .....CTCAAGCACACAGAGTGAGCTCCAGAGTCAGGTGCGCTAC 333
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 GlnLysGlnValGluLysSerGlnAsnLysAlaLysGlnCysLysGlu 183
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
334 AAGGAGGAGTTTGAGAGAACAGGCGCAAGGTTTCCGCTAGTGGCAGA 383
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 rAlaThrGluAlaGluArgValTyrArgGlnAsnIleGluGlnLeuGlu 200
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
384 CACGCCCGAGCTCCAGAGAATCAAGAAG..... 411
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 rGAlaArgThrGluTrpGluGlnGluHisArgThrCysGluAlaPhe 216
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
411 ..... 411
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217 GlnLeuGlnGluPheAspArgLeuThrIleLeuArgAsnAlaLeuTrpVa 233
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
411 ..... 411
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
233 lHisCysAsnGlnLeuSerMetGlnCysValLysAspGluLeuTyrG 250
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
412 .....ACCCAGGACAGATCAGTAATAATAAATACCATGAGGAGTTG 454
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
250 luGluValArgLeuThrLeuGluGlyCysAspValGluGlyAspIleAsn 266
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
455 ACAAGAGCCGATGGCGCCTAGCGGGCGGAGGCGATGGAGCCAGCGGT 504
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
267 GlyPheIleGlnSerLysSerThrGlyArgGluPro..... 278
```

```

505 CCGGATTCACAGCGCGACAGCTACCGCGGCCCTGGAGCAGCAGCA 554
      |||::: |||::: |||:::|
279 .ProAlaProValProTyrGlnAsnTyrTyrAspArgGluValThrProL 295
      ||| ::::| ::| ::::|
555 GCCTCACCACATCCCGACCACT.....GCCCGG 583
      ::::| ::| ::::|
295 euIlleGlySerProSerlleGlnProSerCysGlyValIleLysArgphe 311
      ::::| ::| ::::|
584 TTACACAGCAGCCCGACGACGCG.....GTGGCCCGAGTCCCTAT 624
      ::::|
312 SerGlyLeuLeuHisGlySerProLysThrThrProSerAlaProAlaAl 328
      ::::|
625 GTGGCTAC.....AAGGAGCCTGCAGCCCCA..... 651
328 aSerThrGluThrLeuThrProThrProGluArgAsnGluLeuValTyrA 345
652 .....G 652
345 laSerlleGluValGlnAlaThrGlnGlyAsnLeuAsnSerSerAlaGln 361
      ::::| ::| ::::|
653 TCTCCATACAGCGCGCCCGCAGGTGT.....GGCGGAAG 690
362 AspTyrArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGluLeuAs 378
      ::::| ::| ::::|
691 CGGTACCGCGCGGTGTATGACTACAGCCCGCCGACGAGGAGGCTCTC 740
      ::::| ::| ::::|
378 pIleSerAlaGlyAspIleLeuAlaValIleLeuGluGlyGluAspGlyT 395
      ::::| ::| ::::|
741 CTTCCAGGACGGGGACACCATCGTCAAGTGCAGCAGATCGACAGCGGT 790
395 rPTrp.....ThrValGluArgAsnGlyGlnArgGlyPheValProGly 409
      ::::| ::| ::::|
791 GGATGTACGGGACGCTGGAGCGCACCGGCGACACGGGGATGCTGCCGGCC 840
410 SerTyrLeuGluLysLeu 415
      ::::| ::|
841 AACTACGTGGAGGCATC 858

```

THIS PAGE BLANK (USPTO)

QY 61 ARKAGQTEMNSLRTSFDLSLKQOTENVGSAHIQALALREELSLSEEFERQKQKYE 120
DB 61 ARKAGQTEMNSLRTSFDLSLKQOTENVGSAHIQALALREELSLSEEFERQKQKYE 120
QY 121 AIMDRVOKSKLSLYKTTMESKKAYDKQCRDADDAEQAFERVSANGHQKQVEKSONKAKQC 180
DB 121 AIMDRVOKSKLSLYKTTMESKKAYDKQCRDADDAEQAFERVSANGHQKQVEKSONKAKQC 180
QY 181 KESATEAERYVRONIQLERARTEWQEHRTTCEAFOLQEFDRLLTILRNALWVHCNQLSM 240
DB 181 KESATEAERYVRONIQLERARTEWQEHRTTCEAFOLQEFDRLLTILRNALWVHCNQLSM 240
QY 241 QCVKDDDELYEVRLLTLEGCDVEGDINGFIQSKSTGREPPAPVQNYVDREVTPLIGSPS 300
DB 241 QCVKDDDELYEVRLLTLEGCDVEGDINGFIQSKSTGREPPAPVQNYVDREVTPLIGSPS 300
QY 301 IQPCSGVIRKFSGLLHGSPKTTSPAPAASTETLTPTPERNELVYASIEVQATQGNLNSA 360
DB 301 IQPCSGVIRKFSGLLHGSPKTTSPAPAASTETLTPTPERNELVYASIEVQATQGNLNSA 360
QY 361 QDYRALYDYTAQNSDELDSAGDILAVILEGEDGWTVERNGQGVPGSYLEKL 415
DB 361 QDYRALYDYTAQNSDELDSAGDILAVILEGEDGWTVERNGQGVPGSYLEKL 415

RESULT 2

US-08-938-830-29
; Sequence 29, Application US/08938830
; Patent No. 6040437
; GENERAL INFORMATION:
; APPLICANT: Lasky, Laurence A.
; APPLICANT: Dowbenko, Donald J.
; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
; TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPs)
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,830
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/798419
; FILING DATE: 07-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1066P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-938-830-29

Query Match 79.7%; Score 1714.5; DB 3; Length 400;
Best Local Similarity 77.9%; Pred. No. 2.7e-150;
Matches 338; Conservative 17; Mismatches 26; Indels 53; Gaps 3;

QY 1 MMAQLQFADFRCDFTAHTGYEVLLQRLLDGKMKDKVEELLQRAQAEERYGKELVQI 60
DB 1 MNPQLQFADFRCDFTAHTGYEVLLQRLLDGKMKDKMEELLQRAQAEERYGKELVQI 60
QY 61 ARKAGQTEMNSLRTSFDLSLKQOTENVGSAHIQALALREELSLSEEFERQKQKYE 118
DB 61 ARKAGQTEMNSLRTSFDLSLKQOTENVGSAHIQALALREELSLSEEFERQKQKYE 120
QY 119 -----YEAIMDRVOKSKLSLYKTTMESKKAYDKQCRDADDAEQAFERV 161
DB 121 AVPRQSDCEVKSWSWEYEAIVMDRVOKSKLSLYKTTMESKKAYDKQCRDADDAEQAFERV 180
QY 162 SANGHOKQVEKSONKAKQCKESATEAERYVRQNIQLERARTEWQEHRTTCEAFOLQEF 221
DB 162 SANGHOKQVEKSONKAKQCKESATEAERYVRQNIQLERARTEWQEHRTTCEAFOLQEF 240
QY 222 DRLLTILRNALWVHCNQLSMQCVKDDDELYEVRLLTLEGCDVEGDINGFIQSKSTGREPPAP 281
DB 222 DRLLTILRNALWVHCNQLSMQCVKDDDELYEVRLLTLEGCDVEGDINGFIQSKSTGREPPAP 298
QY 282 VYQNYVDREVTPLIGSPSIQPCSGVIRKFSGLLHGSPKTTSPAPAASTETLTPTPERNE 341
DB 282 VYQNYVDREVTPLIGSPSIQPCSGVIRKFSGLLHGSPKTTSPAPAASTETLTPTPERNE 330
QY 342 LVYASIEVQATQGNLNSAQQDYRALYDYTAQNSDELDSAGDILAVILEGEDGWTVERN 401
DB 342 LVYASIEVQATQGNLNSAQQDYRALYDYTAQNSDELDSAGDILAVILEGEDGWTVERN 386
QY 402 GQGVPGSYLEKL 415
DB 402 GQGVPGSYLEKL 400

RESULT 3

US-08-938-830-26
; Sequence 26, Application US/08938830
; Patent No. 6040437
; GENERAL INFORMATION:
; APPLICANT: Lasky, Laurence A.
; APPLICANT: Dowbenko, Donald J.
; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
; TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPs)
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,830
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/798419
; FILING DATE: 07-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1066P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids

;
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-938-830-26

Query Match 15.1%; Score 325; DB 3; Length 907;

Best Local Similarity 25.9%; Pred. No. 1.9e-21; Mismatches 163; Indels 28; Gaps 7;

Matches 91; Conservative 70; Indels 163; Gaps 28; Gaps 7;

QY 21 GYEVLLQRLDGRKMDVBEELLRORAAQAEERYCKELVQIARKAGGOTEMNSLRTSPDSL 80

Db 13 GMDALSKTSSLSVLESIDEFYAKRASIEREYASKLQELAAASSADIPEVGSILNNILSM 72

QY 81 KQOTENVGSAHIQIALALRELRS-LEEFREKOEORKKYEAIMDRVOKSKLSYKKTME 139

Db 73 RTETGMAKAHEEVSOQINTELRNKIREYIDQTEQKVVAANAIEELYQKTALEIDLSE 132

QY 140 SKKAYDQKCRDADAQAFERVSANGHOKOVEKSONKAKOCKESATEAERVYQNTIOLE 199

Db 133 KDAEYFSCNKLNSYMQTKMTG-----RELDKYNLKIQAALAVKMDAEYRETNELL 188

QY 200 RARTEWQEHRTCEAFQLOQFDRLTILRNALVHNCNOLSMQCVKDDDELYEVRILTLEGC 259

Db 189 TVIREMDRTEVCDRAFQHIIEYRLFLKTNMAYANIISTACVKDDSECKIRLTLENT 248

QY 260 DVEGDINGTQSKTGREPAPVYQNYDRE-----VTPLGSPSIQSPCGVIKRFSG 313

Db 249 NIDEDITQMIQNEGTGTIP-PLPEFNDYKENGCLNYDIDQLISKAPSYPYSSSRPSASA 307

QY 314 LLHGSPKTPPSAASATETILT-----PT-----PERNELVVASIEVQ 350

Db 308 SLASSPTRSAFRPKTS-ETVSEVSVSPPTSLHSPKVPVNEQVEQVTEVE 358

RESULT 4

US-08-938-830-3

; Sequence 3, Application US/08938830

; Patent No. 6040437

; GENERAL INFORMATION:

; APPLICANT: Lasky, Laurence A.

; APPLICANT: Dowbenko, Donald J.

; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage

; TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPs)

; NUMBER OF SEQUENCES: 73

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 1 DNA Way

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WinPatIn (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/938,830

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/798419

; FILING DATE: 07-FEB-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Dreger, Ginger R.

; REGISTRATION NUMBER: 33,055

; REFERENCE/DOCKET NUMBER: P1066P1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650/225-3216

; TELEFAX: 650/952-9881

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 48 amino acids

;
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-938-830-3

Query Match 11.4%; Score 246; DB 3; Length 48;

Best Local Similarity 97.9%; Pred. No. 3.7e-16;

Matches 46; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 366 LYDTAONSDELDISAGDILAVILEGEDGWWTVVERNGORGFVPGSYL 412

Db 1 LYDTAONSDELDISAGDILAVILXGEDGWWTVVERNGORGFVPGSYL 47

RESULT 5

US-08-942-423-4

; Sequence 4, Application US/08942423

; Patent No. 5891673

; GENERAL INFORMATION:

; APPLICANT: Hashimoto, Yasuhiro

; APPLICANT: Takemoto, Yoshihiro

; TITLE OF INVENTION: Lck Binding Protein

; NUMBER OF SEQUENCES: 68

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Syntex (U.S.A.) Inc.

; STREET: 3401 Hillview Ave.

; CITY: Palo Alto

; STATE: California

; COUNTRY: U.S.A.

; ZIP: 94303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/942,423

; FILING DATE: 01-OCT-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/362,715

; FILING DATE: 23-DEC-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Perles, Rohan

; REGISTRATION NUMBER: 35,752

; REFERENCE/DOCKET NUMBER: 28260

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 852-1698

; TELEFAX: (415) 496-3529

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 546 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FRAGMENT TYPE: N-terminal

; ORGANISM: HSI

; US-08-942-423-4

Query Match 7.8%; Score 168.5; DB 2; Length 546;

Best Local Similarity 22.9%; Pred. No. 2.4e-07;

Matches 101; Conservative 50; Mismatches 153; Indels 137; Gaps 19;

QY 52 RYKELVQIARKA-----GGQTEMNSLRTSFDLSLKQOTENVGSAH-----IQALALRE 100

Db 161 RYGVQADRVKSAVGFDYQKTEKH-----ESQDKSYGGKYGDKKVDKVS-AVG 213

QY 101 ELRSLEEFREKQKQKQK-----KYEAIMDRVQK-----SKLSLYKTKMESKAYDQK- 147

Db 214 EYQKTEKHESQDYKVGFGVQTDRODKCALGWDHQLHESQDYKTFGGKF 273
QY 148 ---CRADDAEAF---ERYSANGHOKQVEK-----SQKAKQKESATEAERV----- 190
Db 274 GVQSERQDSSAVGFDYKERLAKHEPOODYAKGGYGVQKDRNDKXNSTFEVQVPSA 333
QY 191 -----YRONIEQLEARTWEQEHRHTTCEAFQLEQFDRILTILRNALMVH 234
Db 334 YQKTVPIEAVTSKTSIRANFENLAKER---EOEDRKAERAO----- 375
QY 235 CNOLSMQVKKDDLYEVRUTLGGCDVEGDIQSKSTGREGPPAPVPYQNYDR----- 290
Db 376 -----RWAKERQOEAREKLEE-----QARAKKQTPPASPSPQPTEDRPSS 418
QY 291 -----EVTPLIGSPISQPCGVKRFSGLLHGSPKTPPSAPAA-----STETLTPPERN 340
Db 419 PIYDAAPFAEPSYRGS-----EPEPEYSIEAGIPEAGSQOGLTYT---S 462
QY 341 ELVYASIEV---QATQGNLNSSAQDYR---ALYDYTAQNSDELIDISAGDILAVILEGE 392
Db 463 EPVYETTEAPGHYQAEEDTDVGYESDLGITAIALYDQAAAGDDEISFPDDIITNEMID 522
QY 393 DGMWTVERNQORGVPQSYLE 413
Db 523 DGMWGVCKRGYGLFPANYVE 543

RESULT 6
US-08-435-454-4
; Sequence 4, Application US/08435454
; Patent No. 5605830
; GENERAL INFORMATION:
; APPLICANT: Prendergast, George C.
; TITLE OF INVENTION: Murine and Human C-Myc Interacting
; TITLE OF INVENTION: Protein and Uses Therefor
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P O Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/435,454
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST60USA
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 451 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-435-454-4

Query Match 7.0%; Score 150.5; DB 1; Length 451;
Best local Similarity 21.7%; Pred. No. 8.1e-06;
Matches 100; Conservative 70; Mismatches 186; Indels 105; Gaps 20;

QY 33 RKMCKDVBEELLRQAAQERYKELVQIARKAGGO--TE-----MNSLRTSFDLSLKQOTENW 87
Db 17 KKLTRAQAEKVLQKLGKADETKDEQFEQCVONENKQLTEGRLQKDLTYLASVKAMHE-- 74
QY 88 GSAHIQALALAPELSLSEEFREKQKRYEAI-----MDRVQSKSLSYKKTME 139
Db 75 -----ASKKLNCEQLQEVVEPDWPGRDEANKIAENNDLLMMDYHOK-----LYDQALL 121
QY 140 SKKAY-----DOKCRDADDAEQAFERVSANGHOKQVE-----KSONKAKQCKESATEAERV 190
Db 122 TMDTYLGQFPDIKSRJAKRGKLVQDYDSARHHVYESJQAKKDEAKIAKAEELIKAKQV 181
QY 191 YRO-NTEQLEARTWEQEHRHTTCEAFQLE-----QEFDRLTILRNALWVHCNO--- 237
Db 182 FEWNVDLQELFSLNSRVGVYVNTFQSTAGLEENFHKEMSKLNQNLNDVLGLEKQHG 241
QY 238 -----ISMQCVKDDLYEVRUTLGGCDVEGDIQSKSTGREGPPAPVPYQNYDR 292
Db 242 SNTFTVKAQPRKSKSLFSLRRKKNSDNAPAKGN---KSPSPDGSPAATP-EIRVNHEP 297
QY 293 TPLIG-----SPS-----IQPSGVKRFSGLLHGSPKTPPSAPAASTET 332
Db 298 EPAGGATPGATLKPSPQPAEASEVAGTQPAAGA--QEPGETSASEAASSSLPAVVVET 355
QY 333 LTPPERNELVYASIEVQATQGNLN---SSAQDYRALYDYTAQNSDELIDISAGDILAVI- 388
Db 356 FPAT-----VNGTVEGSGAGRLDPLPGFMFKYQAOHDYATATDELQKAGGVVLVIP 409
QY 389 -----LEGEDGW-----WTVERNGO--RGVFPQSYLEKL 415
Db 410 FQNPEDQEGWLMGVKESDQWQHKKLEKRCGVFPENETERV 450

RESULT 7
US-08-919-145-6
; Sequence 6, Application US/08919145
; Patent No. 5958753
; GENERAL INFORMATION:
; APPLICANT: Prendergast, George C.
; TITLE OF INVENTION: Bau, A Binl Interacting Protein, and
; TITLE OF INVENTION: Uses Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,145
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/025,482
; FILING DATE: 29-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kodroff, Cathy A.
; REGISTRATION NUMBER: 33,980
; REFERENCE/DOCKET NUMBER: WST73AUSA
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 451 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-919-145-6

Query Match 7.0%; Score 150.5; DB 2; Length 451;
Best Local Similarity 21.7%; Pred. No. 8.1e-06;
Matches 100; Conservative 70; Mismatches 186; Indels 105; Gaps 20;

```
QY 33 RKMCKDVEELLRQRAQAEERYGKELVQIARKAGGQ-TE-----MNSLRTSFDSLKQQTENV 87
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17 KKLTRAQEKVLQKLGKADETKDEQFEQCVQNFNKLQTEGRLQKDLRTYLASVKAMHE-- 74
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 88 GSAHIQALALREELRSLSEEFREKQKRYEAI-----MDRVOKSKLSLYKKTME 139
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 75 -----ASKKLNCEQLQEVYEPDPGRDEANKIAENNDLLWMDYHOK-----LVDQALL 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 140 SKKAY-----DOKCRDADDAEQAFERSVANGHOKVE-----KSNKAKOKESATEAERV 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 TMDTYLGQFPDIKRIAKRGRKLVYDSARHHYESLQTAKKKDEAKIAKAEELIKAQKV 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 191 YRQ-NIEQLERARTEWEQHRTTCEAFQL-----QEPDRLTILRNALWVHCNQ--- 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 FEENVDLQEELPSLWNSRVGFYVNTFQSIAGLEENFHKMSKLNQNLNDVLVGLXKHG 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 238 -----LSMQCVKDDELYEEVRLTLEGCDVEGDINGFIQSKSTGREPPAPVPYQNYDREV 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 242 SNTFTVRAQPRKSKLFSRLRRKNSDNAPAKGN---KSPSPDGSPPATP-EIRVNHEP 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 293 TPLIG-----SPS-----IQPCGVKIRFSGLLHGSFKTTPSAPAASTET 332
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 298 EPAGGATPGATLPKSPQAPAEVAGTQPAAGA--QEPGETSASEAASSLPVAVVET 355
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 333 LTPPTPERNELVYASTEVOATQGNLN---SSAQDYRALYDYTAQNSDELDSAGDILAVI- 388
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 356 FPAT-----VNGIVEGSGAGRLDLPFGFMFKVQAQHDYATDTDELQKAGDWLVIP 409
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 389 -----LEGEDGW-----WTVERNGQ--RGFVPGSYLEKL 415
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 410 FONPEQDEGLMGVKESDNQHKKLEKRCRGVFPENFETRV 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 8
PCT-US96-06231A-4
; Sequence 4, Application PC/TUS9606231A
; GENERAL INFORMATION:
; APPLICANT: Wistar Institute of Anatomy & Biology
; TITLE OF INVENTION: Murine and Human Box-Dependent
; TITLE OF INVENTION: Myc-Interacting Protein (BIN1) and Uses Therefor
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P O Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/06231A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/435,454
; FILING DATE: 05-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST60APCT

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 451 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-06231A-4

Query Match 7.0%; Score 150.5; DB 4; Length 451;
Best Local Similarity 21.7%; Pred. No. 8.1e-06;
Matches 100; Conservative 70; Mismatches 186; Indels 105; Gaps 20;

```
QY 33 RKMCKDVEELLRQRAQAEERYGKELVQIARKAGGQ-TE-----MNSLRTSFDSLKQQTENV 87
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17 KKLTRAQEKVLQKLGKADETKDEQFEQCVQNFNKLQTEGRLQKDLRTYLASVKAMHE-- 74
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 88 GSAHIQALALREELRSLSEEFREKQKRYEAI-----MDRVOKSKLSLYKKTME 139
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 75 -----ASKKLNCEQLQEVYEPDPGRDEANKIAENNDLLWMDYHOK-----LVDQALL 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 140 SKKAY-----DOKCRDADDAEQAFERSVANGHOKVE-----KSNKAKOKESATEAERV 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 TMDTYLGQFPDIKRIAKRGRKLVYDSARHHYESLQTAKKKDEAKIAKAEELIKAQKV 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 191 YRQ-NIEQLERARTEWEQHRTTCEAFQL-----QEPDRLTILRNALWVHCNQ--- 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 FEENVDLQEELPSLWNSRVGFYVNTFQSIAGLEENFHKMSKLNQNLNDVLVGLXKHG 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 238 -----LSMQCVKDDELYEEVRLTLEGCDVEGDINGFIQSKSTGREPPAPVPYQNYDREV 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 242 SNTFTVRAQPRKSKLFSRLRRKNSDNAPAKGN---KSPSPDGSPPATP-EIRVNHEP 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 293 TPLIG-----SPS-----IQPCGVKIRFSGLLHGSFKTTPSAPAASTET 332
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 298 EPAGGATPGATLPKSPQAPAEVAGTQPAAGA--QEPGETSASEAASSLPVAVVET 355
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 333 LTPPTPERNELVYASTEVOATQGNLN---SSAQDYRALYDYTAQNSDELDSAGDILAVI- 388
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 356 FPAT-----VNGIVEGSGAGRLDLPFGFMFKVQAQHDYATDTDELQKAGDWLVIP 409
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 9
US-08-652-972A-4
; Sequence 4, Application US/08652972A
; Patent No. 5723581
; GENERAL INFORMATION:
; APPLICANT: Prendergast, George C.
; APPLICANT: Sakamuro, Daitoku
; TITLE OF INVENTION: Murine and Human Box-Dependent
; TITLE OF INVENTION: Myc-Interacting Protein (BIN1) and Uses Therefor
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P O Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/652,972A

```

; FILING DATE: 24-MAY-1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/435,454
; FILING DATE: 05-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST60BUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 451 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-652-972A-4

Query Match 6.9%; Score 149.5; DB 1; Length 451;
Best Local Similarity 21.7%; Pred. No. 1e-05;
Matches 100; Conservative 70; Mismatches 186; Indels 105; Gaps 20;

QY 33 RKMKDVEELLRQRAAEERYGKELVOIARKAGGQ-TE---MNSLRTSFDLSLKQOOTENV 87
DB 17 KLTTRAQEKVLQKLGKADETKDFQECVONFNKQLTEGLTQKDLRTYLASVKAMHE-- 74
QY 88 GSAHIQIALALRELRSLERFRERQKQKYEAI-----MDRVQKSKLSLYKKTME 139
DB 75 -----ASKKLNELQEVYEPDPWGRDEANKIAENNDLLWMDYHQK-----LVDQALL 121
QY 140 SKKAY-----DOKRDADDAQAFERVSANGHOKQVE---KSONKAKOCKESATEAERV 190
DB 122 TMDTYLGQFPDIKSRIAKRGKLVYDSARHHYESLQTAKKKDEAKTAKAEELIKAKV 181
QY 191 YRO-NIEOLERARTEWQEHRTTCEAFQ-----QEFDRLTILRNALVHNCQ--- 237
DB 182 FEEMNDLQELPSLWNSRVGFYVNTFQSIAGLEENFHKEMSKLNQNLNDVLVGLKQHG 241
QY 238 -----LSMQCKVDDELYEVEVRLTLEGCDVEGDINGFTQSKSTGREPPAPVQNYKYREV 292
DB 242 SNTFTVKAQPRKSKLFLSRLRRKKNNDNAPAKGN---KSPSPDPGSPAATP-EIRVNHEP 297
QY 293 TPLIG-----SPS-----IOPSCGVIKRFSGLLHGSPKTTTPSAPAASSTET 332
DB 298 EPAGGATPGATLPKSPSQPAEASEVAGTQPAAGA--QEPGETAASEAASSSLPAVVVET 355
QY 333 LTPTPERNELVYASIEVQATQGNLN---SSAQDYRALYDYTAQNSDELIDISAGDILAVI- 388
DB 356 FPAT-----VNGIVEGGSGAGRLDLPFGFMFKVQAQHDYATDTDTDELQKAGDVVLV 409
QY 389 ----LEGEDGW-----WTVERNGQ--RGFVPGSYLEKL 415
DB 410 FQNPEDQEGVLMGVKESDWNQHKKLEKRCGVFPENFTERV 450

RESULT 10
US-08-870-126-4
; Sequence 4, Application US/08870126
; Patent No. 6048702
; GENERAL INFORMATION:
; APPLICANT: Prendergast, George C.
; APPLICANT: Sakamuro, Daitoku
; TITLE OF INVENTION: Mice and Human Box-Dependent
; TITLE OF INVENTION: MYC-Interacting Protein (Bin1) and Uses Therefor
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P O Box 457
; CITY: Spring House
; STATE: Pennsylvania

```

Patent No. 5891674
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
TITLE OF INVENTION: INSULIN RECEPTOR TYROSINE KINASE SUBSTRATE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,563A
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0323 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 534 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT04
CLONE: 918158
US-08-878-563A-1

Query Match 6.9%; Score 148.5; DB 2; Length 534;
Best Local Similarity 21.6%; Pred. No. 1.6e-05;
Matches 95; Conservative 72; Mismatches 177; Indels 95; Gaps 21;
QY 35 MCKDVEELLRQRAAEERYGKELV---QIARKAGQTEMNSLRTSFDLSLKQOTEN----- 86
DB 34 MGKRYKALAGVTYAAKGYFDALVKMGELASESGSKELGVDLQFQMAEVHRQIQNQLDEM 93
QY 87 VGSAGHIALALREELRLSLEEF----RERQEKQKRYEALMDRVQSKLSLYKKTWESK 141
DB 94 LKSPHNELLTQLEQKVELDSYLSAALKKKYQTEQSKGDA-LDKCQAEELKRLKKSQSGSK 152
QY 142 ---KAYDOKCRDAD---DAEQAFERVSANGHQKQVKSQNK-----AKQCKESATEAEV 190
DB 153 NPQKYSKELQYIDAISNKQCELENYVSDGYKTTALTEECRRFCFLVEKQCAVAKNSA-AY 211
QY 191 YRQNIQELERARTWEQHRHTTCFAQLQEFDRLTILRNALVHCNQLSMQCVKDDLEYE 250
DB 212 HSKGELLPOKLPLMOQ-----ACA-----DPSKIPERAV-----QLMQQVANSNG--- 251
QY 251 EVRLTLEGCDVEGDINGFIQSKSTGREPPAPVPYQNYDREVTLPGISPSIQSPSCGVIKR 310
DB 252 ---ATFLSALSASKNLVIDSPIGAK-PLDVP-----PELAPFVGRMSAQESTPIMNG 301
QY 311 FSGLLHSGPKTPPSAP-----AASTETLTPTPERNEL-----VYASI 347
DB 302 VTG-----PDGEDYSPWADRKAQPKSLSPQSQSKLSDSYNTLPYRKSVTPKNSYATT 356
QY 348 EVQATQGNLNSA-----QDYRALYDYTA-QNSDELDTISAGDILAVIL-EGEDGW--W 396

DB 357 ENKTLPRSSSSMAAGLERNGRMVRAIFSHAAGDNTLLSFKEGDLITLLVPEARDGWYH 416
QY 397 TVERNRGORGFVPVGSYLEKL 415
DB 417 ESEKTKMRGWFPPFSYTRVL 435
RESULT 12
US-08-466-390-4
Sequence 4, Application US/08466390
Patent No. 5686562
GENERAL INFORMATION:
APPLICANT: TOURKATLY, GARY
APPLICANT: LIDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,390
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-466-390-4
Query Match 6.9%; Score 147.5; DB 1; Length 2101;
Best Local Similarity 22.1%; Pred. No. 0.00017;
Matches 101; Conservative 63; Mismatches 146; Indels 147; Gaps 22;
QY 19 HTGVEVLRLQRLDGRCKMKDVEELLRQRAAEERYGKELVOIARKAGG-----QTEM 70
DB 1350 HTSQALVSELLPAKHLG---QQLAQEQAARAKHREELQSKQAAGLRAELRAQREL 1406
QY 71 -----NSLRTSFDLSLKQOTENVCSAHIQIALALREELSLSE----- 107
DB 1407 GELIPLRQKVAEQRTAQOLRAEKASYAEOLSLMKKAHGLLA-----EENRGLGERANLGR 1462
QY 108 -FRERQREQ-RKKY-----EALMDRVQSKLSLYK-----KTMES 140
DB 1463 QFLEVLDOAREKYVQELAAVRADAETFLAEVQREAGQSTARELEVMTAKYEGAKVKVLEE 1522
QY 141 KKAYDQ-----KCRDADA-----EQAFERVSANG---HOKOVKSKQNKAKQC 180
DB 1523 RQRFEERQKLTQAQVEELSCKKLADSDQASKVQQKKAQAQGGESQOEAFQOQLNEL 1582
QY 181 KESATEAEVRYRQNIQELERARTWE-----QEHRITTCFAQLQ-----EFDRLT 225
DB 1583 QAQLSQEQAHAHYKLQMEKAKTHYDAKKQKQNOQLQELQSLQENKELRAEERL 1642

QY 226 ILRNALWVHCNLSMOCVAKDELYEVR-LTLEGCDVEGDI-NGFTQSKSTG-----275
Db 1643 -----HELQOAGLTKAEQTCRHILTAQVRSLEAQAHAHQDLGLKGFQVATDA 1692
QY 276 ---REPPAPVQNYD-----REVTP-L-IGS--PSIQPSGCVIKRFSGLLHSGSKPTTP 323
Db 1693 LKSRPQAKPOLDSLSDLSCEGTPLSITSKLPRTQP-----DGTSPV 1738
QY 324 SAPAAS-TETLPTPERNELVYASIEVQATOGNLSS 359
Db 1739 GEPASPIORLPKVESLESLEYFTPIPARSQAPLESS 1775
RESULT 13
US-08-470-950-4
; Sequence 4, Application US/08470950
; Patent No. 5698439
; GENERAL INFORMATION:
; APPLICANT: TOUTKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,950
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-470-950-4

Query Match 6.9%; Score 147.5; DB 1; Length 2101;
Best Local Similarity 22.1%; Pred. No. 0.00017;
Matches 101; Conservative 63; Mismatches 146; Indels 147; Gaps 22;
QY 19 HTGYEVLRLDGRKMKDVLELLRQRAQAERYGKELVQIARKAG-----QTEM 70
Db 1350 HTSTQALVSELLPAKHLC---OQLOAQAAAEKREHELEQSKQAAGGLRAELLRAQREL 1406
QY 71 -----NSLRTSFDLSKQOTENVGSAHIQALALREELSLSE-----107
Db 1407 GELIPLRQKVAEQRTAQOLRAEKASYAEQLSMLKKAHGLLA-----EENRGLGERANLGR 1462
QY 108 -FRERQKEQ-RKKY-----EAIMDRVQKSKLSLYK-----KTMES 140
Db 1463 QFLEVLDAQREKYVQELAAVRAADATRLAEVQREQAQSTARELEVNTAKYGAQKVLSE 1522
QY 141 KKAYDO-----KCRDADDA-----EQAFERVSANG---HQQVEKSONKAKQC 180

Db 1523 RORFOEEROKLTAQVEELSCLKLADSDQASKVQOQKLLKAYQAQGESQQAORFQAQLNEL 1582
QY 181 KESATAEARVYRQNIQLEOLARTENE-----QEHRITTCFAFOLO-----EFDRLT 225
Db 1583 QAQLSQKEQAABHYKLQMEKAKTHYDAKKQOQNOEQLEQLSLEQKQENKELRAEAERLUG 1642
QY 226 ILRNALWVHCNLSMOCVAKDELYEVR-LTLEGCDVEGDI-NGFTQSKSTG-----275
Db 1643 -----HELQOAGLTKAEQTCRHILTAQVRSLEAQAHAHQDLGLKGFQVATDA 1692
QY 276 ---REPPAPVQNYD-----REVTP-L-IGS--PSIQPSGCVIKRFSGLLHSGSKPTTP 323
Db 1693 LKSRPQAKPOLDSLSDLSCEGTPLSITSKLPRTQP-----DGTSPV 1738
QY 324 SAPAAS-TETLPTPERNELVYASIEVQATOGNLSS 359
Db 1739 GEPASPIORLPKVESLESLEYFTPIPARSQAPLESS 1775
RESULT 14
US-08-467-781-4
; Sequence 4, Application US/08467781
; Patent No. 5780596
; GENERAL INFORMATION:
; APPLICANT: TOUTKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,781
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-467-781-4

Query Match 6.9%; Score 147.5; DB 1; Length 2101;
Best Local Similarity 22.1%; Pred. No. 0.00017;
Matches 101; Conservative 63; Mismatches 146; Indels 147; Gaps 22;
QY 19 HTGYEVLRLDGRKMKDVLELLRQRAQAERYGKELVQIARKAG-----QTEM 70
Db 1350 HTSTQALVSELLPAKHLC---OQLOAQAAAEKREHELEQSKQAAGGLRAELLRAQREL 1406
QY 71 -----NSLRTSFDLSKQOTENVGSAHIQALALREELSLSE-----107
Db 1407 GELIPLRQKVAEQRTAQOLRAEKASYAEQLSMLKKAHGLLA-----EENRGLGERANLGR 1462

RESULT 15
US-08-195-487-4
; Sequence 4, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
; APPLICANT: TOKUKATLY, GARAH P
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,487
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,701
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESO, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-195-487-4

	Matches	101;	Conservative	63;	Mismatches	146;	Indels	147;	Gaps	22;
QY	19	HTGYEVLRLQLDRKMKDV	EELLRQRQAQAEERYKELVQI	ARKAGG-----OTEM	70					
		: :	: :	: :	: :					
Db	1350	HTSTQALVSLLPAKHIL	---OQLOAQFAAAERHRELEBQSKQAAGLRAELIRAQREL	1406						
QY	71	-----NSLRTSFDSLKQCTENGVSAHIOLALALRELSLEE-	-----107							
			:	:						
Db	1407	GELIPLRQKVAEODRTAQQLRPAEKASVAEOLSMLKRAHGLLA	---ZENRGLGERANLGR	1462						
QY	108	-FREROKEQ-RKKY-	-----EAIMDRVOKSKLSLYK-	-----KTMS	140					
		: :	: : :	:						
Db	1463	QFLEVELDQAREXYVOELAAVRADAETRLAEVGOREAQSTARELEVMTAKYEGAKVKVLEE	1522							
QY	141	KKAYDQ-----KCRDADDA---	EQAERVVSANG--HQQVYEKSQNKAQC	180						
		: : : : : : : : : : : : :	:							
Db	1523	RQFOBEROKLTIAQVEELSCKKLADSQAASKVQOOQKLKAVQAQGESSQEAQRFQAQUNEL	1582							
QY	181	KESATPAERYVRONIEQLERARTWE-----QEHRTTCEAFQJQ-----EPDRLT	225							
		: : : : : :	: : :	:						
Db	1583	QAOLSQKEQAEEHYKLQMEKAKTHYDAKKQONQELQBLRSLEQLQKENKELRAEAERLG	1642							
QY	226	ILRNALWHCNQLSMOCVKDDLEYEEVR-ITLEGCDVEGDI-NGFIOKSTGT	275							
		: : : : : : : :	:							
Db	1643	-----HELQAGLTKTEAQEQTCRHILTAQVRSLEAQVAHADQOQLDLGRKFQVATDA	1692							
QY	276	--REPPAPVPYNYND-----REVTLT-IGS-PSIQPCSGVIKRFSLGHGSPKTPP	323							
			:	:						
Db	1693	LKSRFEQAQPOLDLSDLSDCSEEGTPLSITSKLPRTQP-----DGTSVP	1738							
QY	324	SAPAAS-TETLTPTPERNELIVYASIEVOATOGNUNSS	359							
Db	1739	GPASPISORLPPKVKSLESLEYTFPTIPARSOAPLESS	1775							
		: : : : : : : :	:							

Search completed: September 15, 2000, 11:07:12
Job time: 3609 sec

THIS PAGE BLANK (USPTO)

OM of: US-09-068-377-1 to: GenEmbl:* out_format : pfs

Date: Sep 15, 2000 7:59 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frames_p2n.model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09068377/runat_14092000.165621.22405/app_query.fasta_1.477
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=4.500
-GAPEXT=0.050 -GAPOP=10.000 -GAPEXT=0.500 -GAPOP=6.000
-GAPEXT=7.000 -GAPOP=10.000 -GAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-ALIGN=45 -LOCAL -OUTFMT=pfs -THR_MAX=100 -THR_MIN=0
-MAXLEN=200000000 -USPR=US09068377.cgn1_1.3705 -NCPU=6
-ICPU=3 -LONLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-068-377-1
Query length: 415
Database: GenEmbl:*
Database sequences: 972840
Database length: 892348106
Search time (sec): 998.980000

score_list:

Sequence	Strd Orig	zScore	EScore	Len	Documentation
gb_ro:MMU87814	+ 2152.00	2754.60	2.5e-145	1692	!
gb_pr4:AF038603	+ 1913.50	2450.09	2.2e-128	1428	!
gb_pr4:HS094778	+ 1877.50	2402.54	1.0e-125	1656	!
gb_pr4:AF038602	+ 1748.00	2238.02	1.5e-116	1382	!
gb_ro:MMU8101	+ 736.50	942.93	2.0e-44	1005	!
gb_pr2:ROB01015	+ 503.50	649.25	4.6e-28	562	!
gb_pr2:SPC011C11	+ 302.50	357.63	8.0e-12	22500	!
gb_pr4:AF128536	+ 301.00	373.36	1.1e-12	3255	!
gb_p11:SPC015	+ 297.00	366.36	2.6e-12	3993	!
gb_p12:SPAC2058	+ 297.00	347.88	2.8e-11	30200	!
gb_ro:AF139492	+ 284.50	356.28	9.5e-12	2079	!
gb_ro:AF128535	+ 283.50	351.01	1.9e-11	3217	!
gb_pr1:AB006628	+ 283.00	351.07	1.9e-11	2979	!
gb_pr1:AB037800	+ 282.50	347.84	2.8e-11	3956	!
gb_ro:AF139494	+ 279.50	349.89	2.2e-11	2073	!
gb_ro:AF139493	+ 274.00	343.36	5.0e-11	1956	!
gb_ro:AF104402	+ 273.50	344.57	4.3e-11	1597	!
gb_ro:AF139495	+ 269.00	336.97	1.1e-10	1950	!
gb_ov:GDPS2MR	+ 266.50	329.99	2.8e-10	2947	!
gb_ro:MH74GENE	+ 265.50	325.71	4.8e-10	4091	!
gb_pr4:AF130979	+ 264.00	330.76	2.5e-10	1907	!
gb_p11:AK000577	+ 250.50	315.27	1.8e-09	1559	!
gb_pr2:HS044876	+ 240.50	298.40	1.6e-08	2427	!
gb_ro:AF133184	+ 240.50	287.53	6.4e-08	7979	!
gb_ro:MMU19860	+ 236.50	297.80	1.7e-08	1478	!
gb_p11:AK002203	+ 228.00	285.97	7.8e-08	1634	!
gb_ro:RNAJ3148	+ 227.00	280.38	1.6e-07	2618	!
gb_ro:RNOJ31902	+ 227.00	272.21	4.6e-07	6412	!
gb_in1:ECCANTA	+ 207.00	257.65	3.0e-06	1900	!
gb_in1:ECCG13A	+ 207.00	257.65	3.0e-06	1900	!
gb_ro:AF132672	+ 206.00	250.01	7.9e-06	3812	!
gb_ro:AF127798	+ 206.00	249.51	8.4e-06	4025	!
gb_pr3:AF064243	+ 202.00	241.99	2.2e-05	5287	!
gb_pr4:AF114488	+ 202.00	241.73	2.3e-05	5381	!
gb_pr4:AF114487	+ 202.00	240.09	2.8e-05	6439	!
gb_pr3:AF064244	+ 202.00	239.01	3.2e-05	7247	!
gb_in2:AF054612	+ 199.50	241.82	2.3e-05	3750	!
gb_in2:AF053957	+ 199.50	241.23	2.4e-05	3999	!
gb_ro:AF132478	+ 198.00	239.96	2.9e-05	3723	!
gb_ro:AF132481	+ 198.00	237.00	4.2e-05	5145	!

gb_bt98:AC010581 + 196.50 206.49 0.0021 117796 ! AC010581 Drosophila melan
gb_pr5:AF248540 + 195.00 232.01 7.9e-05 5828 ! AF248540 Homo sapiens inter
gb_pr1:AB033082 + 195.00 231.84 8.1e-05 5938 ! AB033082 Homo sapiens mRNA
gb_ov:AF032118 + 194.50 234.58 5.7e-05 4103 ! AF032118 Xenopus laevis int
seq_name: gb_ro:MMU87814
seq_documentation_block:
LOCUS MMU87814 1692 bp mRNA ROD 04-MAR-1997
DEFINITION Mus musculus PEST phosphatase interacting protein mRNA, complete cds.
ACCESSION U87814
VERSION U87814.1 GI:1857711
KEYWORDS house mouse.
SOURCE
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 1692)
Lasky, L.
TITLE PEST Phosphatase Interacting Protein (PSTPIP)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1692)
Lasky, L.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1997) Molecular Oncology, Genentech, 460 Pt. San Bruno Blvd., San Francisco, CA 94080, USA
FEATURES
1..1692
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_type="hematopoietic progenitor cells"
274..1521
/note="similar to S. pombe Cdc15p, a protein associated with formation of the cleavage furrow during cytokinesis; actin interacting protein with C-terminal SH3 domain; PSTPIP"
/codon_start=1
/product="PEST phosphatase interacting protein"
/protein_id="AAB48483.1"
/db_xref="GI:1857711"
/translation="MMAQLQFRDAFWCRDFTAHYGYEVLLQRLLDGRKMKDKVRELLR
RSLAEERYKELVOIARKAGQTEMNSLRSTFDSLKQOTENVGSAHQLALAREEL
RSLEEFERKEQKKYEAMIDRVQKSLSLYKKTMSKKAYDKQKRDADAEQAFOL
VSANGHOKVQKSNQCKESAEVRYQNIQLEERAEVQEHRTTCGAFOL
QEEDRLTLNALVHNCOLSMQCVDDELYEVRLLLEGCDVEDINGFTQSKSTGR
EPAPVPYQNYDREVTPPLGSPSIQPCSGVIKRPSGLLHSGSPKTTTSAPAASTETLT
PTPERNELVTASIEVQATQGNLSSAODYRALYDITQNSDELDSAGDILAVILEGE
DGWMTVERNGQRFVPGSYLEKL"
BASE COUNT 409 a 443 c 526 g 314 t
ORIGIN
alignment_scores:
Quality: 2152.00 Length: 415
Ratio: 5.186 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-068-377-1 x MMU87814 ..
Align seg 1/1 to: MMU87814 from: 1 to: 1692
1 MetMetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPhe 17
|||||
274 ATGATGGCCCACTGCAGTCCGAGATGCTTCTGTGCGAGGACTTCAC 323
17 rAlaHisThrGlyTyrGluValLeuLeuGlnArgLeuLeuAspGlyArgL 34
|||||
324 GCGCCACACAGGGTATGAGTCTACTGCGAGGCTGCTGGACGGCAGGA 373
34 ysMetCysLysAspValGluGluLeuLeuArgGlnArgAlaGlnAlaGlu 50
|||||

```
374 AGATGTGCAAGGATGTGGAGGAGTGTCTACAGACAGAGGGCCAGCGGAG 423
51 GluArgTyrGlyLysGluLeuValGlnIleAlaArgLysAlaGlyGlyG1 67
|||||
424 CAGAGGTACGGGAAGGAGCTGCTCAGATTCACGCAAGGCTGGTGCCA 473
|||||
67 nThrGluMetAsnSerLeuArgThrSerPheAspSerLeuLysGlnInt 84
|||||
474 GACAGAGATGAATTCCTTGAGCAGCTCCTTTGACTCCCTGAAGCAGCAA 523
|||||
84 hrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
|||||
524 CAGAGATGTGGCAGTGCACATCCAGCTGGCCCTGGCCCTCGCTGAG 573
|||||
101 GluLeuArgSerLeuGluGluPheArgGluArgGlnLysGluGlnArgLy 117
|||||
574 GAGCTGGCGGAGCTGGAGGAGTCCGAGAGAGACAGAAAGACCGCGAA 623
|||||
117 sLysTyrGluAlaIleMetAspArgValGlnLysSerLysLeuSerLeuT 134
|||||
624 GAAGTATGAGGCCATCATGGACCGCTGCCAAGACGAAGTTGTGCTCT 673
|||||
134 yrLysLysThrMetGluSerLysLysAlaTyrAspGlnLysCysArgAsp 150
|||||
674 ACAAGAGACCATGGATGCCAAGAGGCATATGACCAGAGTGCAGGGAT 723
|||||
151 AlaAspAspAlaGluGlnAlaPheGluArgValSerAlaAsnGlyHisG1 167
|||||
724 GCACATGATGCTGAGCAGCGCTTCGAGCGGTGTGAGTCCCAATGCCACCA 773
|||||
167 nLysGlnValGluLysSerGlnAsnLysAlaLysGlnCysLysGluSerA 184
|||||
774 GAACCAAGTAGAAAAGACCCAGAACAAAGCCACAGTGCAGAGGAGTCAG 823
|||||
184 laThrGluAlaGluArgValTyrArgGlnAsnIleGluGlnLeuGluArg 200
|||||
824 CCACAGAGGACGAAAGAGTGTACAGGCAAAATATCGAACCACTGGAGAGA 873
|||||
201 AlaArgThrGluLysGluGlnGluHisArgThrThrCysGluAlaPheG1 217
|||||
874 GCGAGGACCGAGTGGGAGCAGGACCGGACTACCTGTGAGGCGCTTCCA 923
|||||
217 nLeuGlnGluPheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValH 234
|||||
924 GTTCAGAGAGTTTGACCGCGCTCACCATCTCCGCAATGCCCTGTGGTGC 973
|||||
234 iCysAsnGlnLeuSerMetGlnCysValLysAspAspGluLeuTyrGlu 250
|||||
974 ACTGTACACAGCTCTCCATGCAGTGTCAAGGATGATGAGCTCTATGAG 1023
|||||
251 GluValArgLeuThrLeuGluGlyCysAspValGluGlyAspIleAsnG1 267
|||||
1024 GAAGTGGCGGTGACCCCTTGAGGGCTGTGATCGGAAGTGCACATCATGG 1073
|||||
267 yPheIleGlnSerLysSerThrGlyArgGluProProAlaProValProT 284
|||||
1074 CTTTCATCCAGTCCCAAGACACTGCAGACAGAGCCGCCAGCTCCGCGTCTT 1123
|||||
284 yrGlnAsnTyrTyrAspArgGluValThrProLeuIleGlySerProSer 300
|||||
1124 ATCAGAACTACTATGACAGGAGGTGACCCCACTGATGGCAGCCCTAGC 1173
|||||
301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisG1 317
|||||
1174 ATCCAGCCCTCTCGCGTGTGATAAAGAGGTCTCTCGGCTGTACATGG 1223
|||||
317 ySerProLysThrThrProSerAlaProAlaAlaSerThrGluThrLeuT 334
|||||
1224 AAGTCCCAAGCACCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1273
|||||
334 hrProThrProGluArgAsnGluLeuValTyrAlaSerIleGluValGln 350
|||||
1274 CTCCCACCCCTGAGCGGATGAGTTGGTCTTACGCATCTCCATCGAAGTGCAG 1323
|||||
```

```
351 AlaThrGlnGlyAsnLeuAsnSerSerAlaGlnAspTyrArgAlaLeuTy 367
|||||
1324 GCGACCCAGGAAACCTTAACATCATCAGCCAGCACTACCGGGCAGCTTA 1373
|||||
367 rAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspI 384
|||||
1374 CGACTACACTGCACAGAAATCTGATGAGCTGGACATTTCCGCGGGAGACA 1423
|||||
384 leLeuAlaValIleLeuGluGlyGluAspGlyTyrTrpThrValGluArg 400
|||||
1424 TCCTGGCGGTCACTCTGGAAGGGGAGGATGGCTGGTGGACTGTGGAGCG 1473
|||||
401 AsnGlyGlnArgGlyPheValProGlySerTyrLeuGluLysLeu 415
|||||
1474 AACGGACACAGTGGCTTTGTCCCTGGGTGCTACTTGGAGAAGCTC 1518
|||||
seq_name: gb_pr4:AF038603
seq_documentation_block:
LOCUS AF038603 1428 bp mRNA PRI 02-FEB-1999
DEFINITION Homo sapiens CD2 binding protein 1 long form mRNA, complete cds.
ACCESSION AF038603
VERSION AF038603.1 GI:2921550
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1428)
AUTHORS Li,J., Nishizawa,K., An,W., Hussey,R.E., Lialios,F.E., Salgia,R.,
Sunder-Plassmann,R. and Reinherz,E.L.
TITLE A cdc15-like adaptor protein (CD2BP1) interacts with the CD2
cytoplasmic domain and regulates CD2-triggered adhesion
EMBO J 17 (24), 7320-7336 (1998)
JOURNAL 99077800
MEDLINE
AUTHORS Li,J., Nishizawa,K., An,W., Hussey,R.E., Lialios,F.E.,
Sunder-Plassmann,R. and Reinherz,E.L.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-1997) Immunobiology, Dana-Farber Cancer
Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES
Location/Qualifiers
1..1428
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="T cell"
1..1251
/note="CD2BP1"
/codon_start=1
/product="CD2 binding protein 1 long form"
/protein_id="AAD11959.1"
/db_xref="GI:2921551"
/translation="MMPQLQFDKFWCRDFTTAHTGYEVLLQLRLDGRMKCKDMEELLR
QRAQAEYRGKELVQIARAGGQTEINSLRASFDLSKQOMENVGSHIQALTLREEL
RSLEEFREQRQKRYEAVDRVQKSLSLYKKAMESKTYEQKCRDADDAEQAFER
ISANGHQVEKSONKARCKDSATEAEVYRQSLAQLEKVAEWEQGHRTTCEAFOL
QEDRLTLRNALVHNSLMOCVKDDDELYEEVRLTLEGCSIDADIDSFIOAKSTGT
EPAPVPYONYDREVTPLTSSPGIOPSCGMKRPESGLHSGSPKTTSLAASASTETL
TPTPERNEGVTATAVOEQINPASPAPAEYRALDYDTAQNDELQSLDGLILEVLEG
EDGWTVERNGORGFVPGVLEKL"
BASE COUNT 352 a 396 c 462 g 218 t
ORIGIN
```

```
alignment_scores:
Quality: 1913.50 Length: 416
Ratio: 4.808 Gaps: 1
Percent Similarity: 95.673 Percent Identity: 87.981
alignment_block:
US-09-068-377-1 x AF038603
```


Align seg 1/1 to: AF038603 from: 1 to: 1428

```
1 MetMetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheTh 17
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1 ATGATGCCCCAGCTGCAGTTCAAAGATGCTTTTGGTGCAGGAGCTTCAC 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
17 rAlaHisThrGlyTyrGluValLeuLeuGlnArgLeuLeuAspGlyArgL 34
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
51 AGCCACACAGCGGCTACGAGGTGCTGCTGAGCGGCTTCTGGATGGCAGGA 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
34 yMetCysLysAspValGluGluLeuLeuArgGlnArgAlaGlnAlaGlu 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
101 AGATGTGCAAGACATGGAGGAGCTACTGAGGCAGAGGCCGCCAGGCGGAG 150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
51 GluArgTyrGlyLysGluLeuValGlnIleAlaArgLysAlaGlyGlyG 67
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
151 GAGCGGTACGGGAAGGAGCTGTGTGTCAGATCGCACGGAAGCAGGTGGCCA 200
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
67 nThrGluMetAsnSerLeuArgThrSerPheAspSerLeuLysGlnGlnT 84
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
201 GACGGAGATCAACTCCCTGAGGCGCTCTTGTACTCTTGAAGCAGCAAA 250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
84 hrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaArgGlu 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
251 TGGAGAAATGTGGCAGCTCACATCCAGCTGGCCCTGACCTGCGCTGAG 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
101 GluLeuArgSerLeuGluGluPheArgGluArgGlnLysGluGlnArgLy 117
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
301 GAGCTGCGGAGTCTCGAGGAGTTCTGTGAGAGCAGAGCAGAGGAA 350
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
117 slsYtyrGluAlaIleMetAspArgValGlnLysSerLysLeuSerLeuT 134
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
351 GAAGTATGAGCGGCTCATGACCGGTGCAGAGAGCAGCAAGCTGCTGCT 400
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
134 yLysLysThrMetGluSerLysLysAlaTyrAspGlnLysCysArgAsp 150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
401 ACNAGAGGCCATGGAGTCCAGAGACATACGAGCAGAGTGCCTGGGAC 450
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
151 AlaAspAlaGluGlnAlaPheGluArgValSerAlaAsnGlyHisG 167
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
451 GCGGACGCGCGAGCAGCGCTTCGAGCGCATTAGCGCCCAAGCGGCACCA 500
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
167 nLysGlnValGluLysSerGlnAsnLysAlaLysGlnCysLysGluSerA 184
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
501 GAAGCAGGTGGAGAGAGTTCAGAACAAACCCAGGAGTGCAGGACTCGG 550
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
184 laThrGluAlaGluArgValTyrArgGlnAsnIleGluGlnLeuGluArg 200
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
551 CCACCGAGCAGCAGCGGTATACAGCAGAGCATTCGCGAGCTGGAGAAG 600
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
201 AlaArgThrGluTrpGluGlnGluHisArgThrThrCysGluAlaPheG 217
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
601 GTCCGGGCTGAGTGGGAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCTTCA 650
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
217 nLeuGlnGluPheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValH 234
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
651 GTGCAAGAGTTGTACCGGCTCACCATTCTCCGCAACGCGCTGTGGGTGC 700
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
234 isCysAsnGlnLeuSerMetGlnCysValLysAspAspGluLeuTyrGlu 250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
701 ACAGCAACAGAGCTCTCCATGCAGTGTGTCAAGATGATGAGCTCTACGAG 750
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
251 GluValArgLeuThrLeuGluGlyCysAspValGluGlyAspIleAsnG 267
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
751 GAAATGCGGCTGACGCTGGAAGGCTGCAGCATAGACGCCGCCACATCCACAG 800
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
267 yPheIleGlnSerLysSerThrGlyArgGluProProAlaProValProT 284
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
801 TTTTATCCAGGCCAAGAGCAGCGGCCACAGAGCCCGCTCCGCTGCCCT 850
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
284 yGlnAsnTyrTyrAspArgGluValThrProIleGlySerProSer 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
851 ACCAGAACHATTACGATCGGGAGGTCAACCCGCTGACACGAGCAGCCTGGC 900
```

```
301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisG 317
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
901 ATACAGCCGTCCTCGCGCATATAAGAGGTTCTCTGGACTGTGCACGG 950
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
317 ySerProLysThrThrPro...SerAlaProAlaAlaSerThrGluThrL 333
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
951 AGTCCCAAGACCACTCGTTGGTGGCTTCTGCTGCTCCACAGAGACC 1000
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
333 euThrProThrProGluArgAsnGluLeuValTyrAlaSerIleGluVal 349
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1001 TGACCCCCACCCCGAGCGGAGTGGGTCTCTACAGGCATCGCAGTG 1050
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
350 GlnAlaThrGlnGlyAsnLeuAsnSerSerAlaGlnAspTyrArgAlaLe 366
||| ||||||| ||| ||||||| ||||||| ||||||| |||||||
1051 CAGGAGATACAGGCAACCGCGCTCCACCAGCCAGGAGTACCGGGGCT 1100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
366 uTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyA 383
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1101 CTACGATTATACAGCGCAGAACCCAGATGAGCTGGACCTCGCGGGAG 1150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
383 spIleLeuAlaValIleLeuGluGlyGluAspGlyTyrThrValGlu 399
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1151 ACATCTGGAGGTGATCTCGAAGGGAGGATGGCTGGTGGACTGGGAG 1200
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
400 ArgAsnGlyGlnArgGlyPheValProGlySerTyrLeuGluLysLeu 415
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1201 AGAACGGGCGAGCTGCTTCGTCCTCCCTGCTCTCTACCTCGAGAAGCTT 1248
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

seq_name: gb_pr4:HSU94778

seq_documentation_block:

LOCUS HSU94778 1656 bp mRNA PRI 05-JAN-1999

DEFINITION Human PEST phosphatase interacting protein homolog (H-PIP) mRNA, complete cds.

ACCESSION U94778

VERSION U94778.1 GI:4100161

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 1656)

AUTHORS Wilson,L.A., Fields,D., Cruz,L., Lasky,L., Friesen,J. and

Siminovitch,K.A.

TITLE The human homologue of mouse PTP-PIP interactor protein

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1656)

AUTHORS Wilson,L.A., Fields,D., Cruz,L., Lasky,L., Friesen,J. and

Siminovitch,K.A.

TITLE Direct Submission

JOURNAL Submitted (20-MAR-1997) Samuel Lunenfeld Research Institute, Mount

Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5,

Canada

FEATURES

source

1..1656 Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_type="activated leukocyte"

217..1656

/gene="H-PIP"

217..1467

/gene="H-pip"

/note="similar to Mus musculus PEST phosphatase

interacting protein, encoded by GenBank Accession Number

U87814; encodes potential N-terminal coiled coil domain"

/codon_start=1

/product="PEST phosphatase interacting protein homolog"

/protein_id="AAD00762.1"

/db_xref="GI:4100162"

/translation="MMPQLQKDAFCRDFTAHTCYEVLLQRLLDGRKMKDMEELLR

QRAHAEYRGELVQIARRKAGQEIINSRLASFDSLKQKMENVGSSHLQLALTREEL

RSLEFRERQEKRRKYEAVMDRVOKSLSLYKKAWESKKTVEHKRLDSDADHAER

[illegible]

AUTHORS
Li, J., Nishi
Sunder-Plasse

AUTHORS
Li, J., Nishizawa, K., A
Sunder-plasman P. and

```
TITLE Direct Submission
JOURNAL Submitted (15-DEC-1997) Immunobiology, Dana-Farber Cancer
Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES
source      Location/Qualifiers
1..1382
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="T cell"
1..1194
/note="CD2BPIs"
/codon_start=1
/product="CD2 binding protein 1 short form"
/protein_id="A011958.1"
/db_xref="GI:2921549"
/translation="MNPQLQFKDAFWCRDFTAHTGYEVLQIRLLDGRKMKDMEEELLR
QRAQAEERQKELVQLARKAGGOTEINSLRASPDLSKQQMENVGSSHIQLALFLREL
LSLEPERQKQKRYEAVMDRVOKSLSLYKKAMESKTYEOKRDADDAQAFER
ISANGHQKQVEKSONKAROCKDSATEAEVYRQSIQALEKVRAEWQEHRTTCEAFQL
QEPDRLLILRNALWVHSNQLSMQCVKDDLEYERVLTLLEGCSIDADIDSFIAKSTGT
EPGEVRLADSAASRFSGLLHSPKTSLSAASASTETLTPTPERNEGVTALAVQEI
QGNPSPAQEYRALYDYTAQNPDLDLSAGDILLEVILEGDEGWMVTVERNGQRFVPGS
YLEKL"
BASE COUNT      344 a 375 c 453 g 210 t
ORIGIN
alignment_scores:
  Quality: 1748.00      Length: 416
  Ratio: 4.661          Gaps: 2
Percent Similarity: 90.144 Percent Identity: 81.971
alignment_block:
US-09-068-377-1 x AF038602 ..
Align seg 1/1 to: AF038602 from: 1 to: 1382
1 MetMetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheTh 17
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1 ATGATGCCCCAGCTGCAGTCTCAAGATGCCTTTTGGTCAGGGGACTTCAC 50
17 rAlaHisThrGlyTyrGluValLeuLeuGlnArgLeuLeuAspGlyArgL 34
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
51 AGCCNCACGGGCTACGAGGTGCTGCTGCACGGGCTTCTGGATGCCAGGA 100
34 ysMetCysLysAspValGluGluLeuLeuArgGlnArgAlaGlnAlaGlu 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
101 AGATCTGCAAGACATGAGGAGCTACTAGGCGACAGGGCCACGGCGGAG 150
51 GluArgTyrGlyLysGluLeuValGlnIleAlaArgLysAlaGlyGlyG 67
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
151 GAGCGGTACGGGAAGGAGCTGGTCAGATCGCACGGAAGGAGGTGGCCA 200
67 mThrGluMetAsnSerLeuArgThrSerPheAspSerLeuLysGlnGln 84
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
201 GACGGAGATCAACTCTTCAGGGGCTCTTTGACTCTCTTGAAGCAGCAAA 250
84 hrGluAsnValCysSerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
251 TGGAGAAATGCGGAGCTCACACATCCAGCTGGCCCTGACCCCTGGCTGAG 300
101 GluLeuArgSerLeuGluGluPheArgGluArgGlnLysGluGlnArgGly 117
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
301 GAGCTGCGGAGTCTCGAGAGTCTTCTGAGAGGCGAAGAGGAGGAGAA 350
117 sLysTyrGluAlaIleMetAspArgValGlnLysSerLysLeuSerLeuT 134
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
351 GAAGTATGAGCGGTCATGGACCGGGTCAGAAAGACAGAGCTGTCGCTCT 400
134 yrLysLysThrMetGluSerLysLysAlaTyrAspGlnLysCysArgAsp 150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
401 ACAAGAAGCCCATGAGTCCCAAGAGACATACGACGACGAAGTCCGGGAC 450
151 AlaAspAspAlaGluGlnAlaPheGluArgValSerAlaAsnGlyHisG 167
```

```
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
451 GCGGACGACGGGAGCAGCGCTTCGAGCGCATTAGCGCAACGGCCACA 500
167 nLysGlnValGluLysSerGlnAsnLysAlaLysGlnCysLysGluSerA 184
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
501 GNAAGCAGGTGGAGAAGAGTCAGAACAAAGCCAGGAGTCGCAAGGACTCG 550
184 laThrGluAlaGluArgValTyrArgGlnAsnIleGluGlnLeuGluArg 200
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
551 CCACCGAGCAGCAGCGGTATACAGGCACAGCATTCGCGCAGCTGGAGAAG 600
201 AlaArgThrGluTrpGluGlnGluHisArgThrThrCysGluAlaPheG 217
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
601 GTCCGGGTGAGTGGGACGAGGACCGGACCACCTGTGAGGCGCTTCA 650
217 nLeuGlnGluPheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValH 234
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
651 GCTGCAAGAGTTGACCGGCTGACCATTCCTCGCAACGCCCTGTGGGTGC 700
234 lsCysAsnGlnLeuSerMetGlnCysValLysAspGluLeuTyrGlu 250
|| ||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
701 ACAGCAACCCAGCTCTCCATGCAGTGTGTCAAGGATGATGAGCTCTACGAG 750
251 GluValArgLeuThrLeuGluGlyCysaspValGluGlyAspIleAsnG 267
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
751 GAAGTGCGGCTGACGCTGGAAGGCTGCAGCATAGACGCGGACATCGACAG 800
267 yPheIleGlnSerLysSerThrGlyArgGluProAlaProAlaProValP 284
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
801 TTTCAATCCAGGCCAAGACGCGGACAGAGCCCGGCTGAGGTCCGCGC 850
284 yrGlnAsnTyrTyrAspArgGluValThrProLeuIleGlySerProSer 300
:::
851 TTGCGGAC..... 858
301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisG 317
::: ||||| ||||||| ||||||| ||||||| ||||||| |||||||
859 .....AGCGACGCTCTAGGTTCTCTGGACTGCTGCACGG 893
317 ySerProLysThrThrPro...SerAlaProAlaAlaSerThrGluThrL 333
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
894 RAGTCCCAAGACCACTTCGTTGGCAGCTTCTCTGGTCCACAGAGACCC 943
333 euThrProThrProGluArgAsnGluLeuValTyrAlaSerIleGluVal 349
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
944 TGACCCCAACCCCGAGCGGAATGAGGTGTCTACACAGCCATCCAGTG 993
350 GlnAlaThrGlnGlyAsnLeuAsnSerSerAlaGlnAspTyrArgAlaLe 366
||| ||||||| ||| ||||||| ||| ||||||| ||| |||||||
994 CAGGAGATACAGGGAACCCCGCCTCACCCAGCCAGGAGTACCGGGCGCT 1043
366 uTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyA 383
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1044 CTACCATATACAGCGCAGACCCAGATGAGCTGACCTGTCGCGGGAG 1093
383 spIleLeuAlaValIleLeuGluGlyGluAspGlyTrpTrpThrValGlu 399
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1094 ACATCTTGAGGTGATCTTGAAGGGAGGAGTGGCTGGTGGACTGTGGAG 1143
400 ArgAsnGlyGlnArgGlyPheValProGlySerTyrLeuGluLysLeu 415
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1144 AGGAACGGGACGCTGGCTTCGCTCCCTGCTTCTACCTGGAGAAGCTT 1191
seq_name: gb_ro:NMU18101
seq_documentation_block:
LOCUS NMU18101 1005 bp mRNA ROD 30-NOV-1998
DEFINITION Mus musculus mRNA for macrophage
actin-associated-tyrosine-phosphorylated protein.
ACCESSION Y18101
VERSION Y18101.1 GI:3947711
KEYWORDS macrophage actin-associated-tyrosine-phosphorylated protein; mayp
gene.
```

```

SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1005)
AUTHORS    Yeung,Y.G., Soldera,S. and Stanley,E.R.
TITLE      A novel macrophage actin-associated protein (MAYP) is
            tyrosine-phosphorylated following colony stimulating factor-1
            stimulation
JOURNAL     J. Biol. Chem. 273 (46), 30638-30642 (1998)
MEDLINE     99023997
REFERENCE   2 (bases 1 to 1005)
AUTHORS    Soldera,S.
TITLE      Direct Submission
JOURNAL     Submitted (10-SEP-1998) S. Soldera, Albert Einstein College of
            Medicine, 1300, Morris Park Avenue, Bronx, New York 10461, USA
FEATURES
    source
        1..1005
            /organism="Mus musculus"
            /strain="(BALB/c x A.CA)F1"
            /db_xref="taxon:10090"
            /cell_line="BAC1.2F5"
        1..1005
            /gene="mayp"
        1..1005
            /gene="mayp"
            /codon_start=1
            /product="macrophage
            actin-associated-tyrosine-phosphorylated protein"
            /protein_id="CAA77027.1"
            /db_xref="GI:3947712"
            /translation="MTGSLFKGNFWSIDILSTIGVDSIIQHLNNGRKNCKEFDFLKE
            RASIEEKYKDLLNLSRKXKPGQSEINTLKRALVEFKQVDNVACHQIQAOTLREEA
            RMEEFRKQKLORKKTTIMDAHKORNAOFKFKAMDAKNVQKCRDKDEAEQAVHR
            SANVANQOEKLFVKLATSKTAVEDSKAYMLHNLMEKVEDWQSHIKACEVFEA
            QBCERINFRNALWHLNQLSQOCVANDMEQVRKSLSEIKDIQYFVNQRATGQ
            TPPAPIMENFYSPORNAAPGKTTGPNPARRGPUPVKRIPDDPDYSVVEDYSLLYQ
            "
BASE COUNT 325 a 227 c 259 g 194 t

alignment_scores:
    Quality: 736.50      Length: 364
    Ratio: 2.888         Gaps: 5
    Percent Similarity: 70.055      Percent Identity: 40.110

alignment_block:
US-09-068-377-1 x MMU18101
..
Align seg 1/1 to: MMU18101 from: 1 to: 1005

7  PheArgAspAlaPheTrpCysArgAspPheThrAlaHisThrGlyTyrG1 23
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16  TTCAAGGGGAACCTCTGGACGACGGACATTCTCAGCACCATTTGGCTACGA 65
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23  uValLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValG 40
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66  CAGCATTTCCAGCACCTCGAACAATGSCCGCTGAAGAAGTGCAGAGTTTG 115
40  luGluLeuLeuArgGlnArgAlaGlnAlaGluGluArgTyrGlyLysGlu 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116  AAGACTTTTGAAGANAGAGACATCAATTGAAGAGAAATATGGCAAGAC 165
57  LeuValGlnIleAlaArgLysAlaGly...GlyGlnThrGluMetAsnse 72
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166  CTGCTGAACCTCTCCAGGAAGACCGCTGGACAGTCTGAGATCAATAC 215
72  rLeuArgThrSerPheAspSerLeuLysGlnGlnThrGluAsnValGlyS 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
216  CTTGAAGAGAGCCCTTGAGTCTCTCAAGCAGCAAGTAGACAACGTGGCTC 265
89  erAlaHisIleGlnLeuAlaLeuAlaLeuArgGluGluLeuArgSerLeu 105

```

```

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
266  AATGTCATTCAGCTTGCAGACTCTAAGAGAAGAGCGGCAGGAAGATG 315
106  GluCluPheArgGluArgGlnLysGluGlnArgLysLysTyrGluAlaI1 122
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
316  GAGAAATTCGGGAAAGACAGAGCTCGCGGAAAGACAGAAACAAT 365
122  eMetAspArgValGlnLysSerLysLeuSerLeuTyrLysThrMetG 139
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
366  ANTGGATGTCGCCACAAGCAAGGAACGCACAGTTCAGANAGCCATGG 415
139  luSerLysAlaTyrAspGlnLysCysArgAspAlaAspAlaGlu 155
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
416  ATGCCAAGAAGAATTATGACAAAGTCCGGGACAAGGATGAGGCAGAG 465
156  GlnAlaPheGluArgValSerAlaAsnGlyHisGlnLysGlnValGluL 172
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
466  CAAAGCTGTCACCGCAGCGCAATGTGGCCCAACCAACGCAACAGAA 515
172  sSerGlnAsnLysAlaLysGlnCysLysGluSerAlaThrGluAlaGlu 189
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
516  GCATTTGTGAAACTGCCCACTTCAAGACTGCAGTAGAGATTTCAGACA 565
189  rgValTyrArgGlnAsnIleGluGlnLeuGluArgAlaThrGluTrp 205
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
566  AAGCGTACATGTTGCACATCAACATGCTGGAGAAGGTTTCGAGAAGACT 615
206  GluGlnGluHisArgThrThrCysGluAlaPheGlnLeuGlnPheAs 222
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
616  CAGAGTGAACACATTAAGCGCTCGAGGTGTTCCGAGCTCAGGAATGTGA 665
222  pArgLeuThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeu 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
666  ACGNATCAACTTCTCCGGAATGCACATGCTGTTCCATCTCAATCAGCTGT 715
239  erMetGlnCysValLysAspGluLeuTyrGluGluValArgLeuThr 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
716  CACAACAATGTGCGAAATGACGAGATGTATGAGCAAGTCCGTAAGAGT 765
256  LeuGluGlyCysAspValGluGlyAspIleAsnGlyPheIleGlnSerLy 272
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
766  TTAGAACATGCGACATGTAGAAGGACATCCAGTATTTGTGAACCCAGCG 815
272  sSerThrGlyArgGluProAlaProValProTyrGlnAsnTyrTrA 289
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
816  GAAACTGGACAGACTCCCCCAGCACCATCATGTATGAGAACTTCTAC. 864
289  spArgGluValThrProLeuIleGlySerProSerIleGlnProSerCys 305
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
865  .....TCTCCTCAG..... 873
306  GlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLysThrTh 322
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
874  .....CGGAATGCTGCCCCACCAGGAAA 896
322  rProSerAlaProAlaAlaSerThrGluThrLeuThrProThrProGluA 339
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
897  GACTACAGGGCCCAATCCAGCAAGGAGAGACCTCTCCCTGTCCCTAAGC 946
339  rg.....AsnGluLeuValTyrAlaSerIleGluValGlnAlaThrGln 353
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
947  GGATACCATGATGCTGATTACTCTGTGTTGAA..... 981
354  GlyAsnLeuAsnSerSerAlaGlnAspTyrArgAlaLeuTyr 367
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
982  .....GATTACAGTTTGTCTCTAT 999

```

seq_name: gb_pr2:IROEST015

seq_documentation_block:

LOCUS IROEST015 562 bp mRNA PRI

DEFINITION Homo sapiens EST from clone 123055, 5' end.

ACCESSION AL110110

```

VERSION      AL110110.1  GI:5791558
KEYWORDS     EST - expressed sequence tag.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 562)
AUTHORS      Auffray,C., Ansorge,W., Ballabio,A., Estivill,X., Gibson,K.,
             Lehrach,H., Poustka,A. and Lundeberg,J.
TITLE        The European IMAGE consortium for integrated Molecular analysis of
             human gene transcripts
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 562)
             Carim,L., Estivill,X., Sumoy,L. and Escarceller,M.
TITLE        Direct Submission
JOURNAL      Submitted (25-AUG-1999) Dept. Genetica Molecular, Institut de
             Recerca Oncologica (IRO), Hospital Duran i Reynals, Autovia de
             Castelldefels Km 2,7 L'Hospitalet de Llobregat, 08907 Barcelona,
             Catalunya, SPAIN. Tel: +34-93-260-7775 Fax: +34-93-260-7776 WWW
             site: http://www.iro.es e-mail enquiries: lsumoy@iro.es,
             mescarceller@iro.es
COMMENT      EURO-IMAGE Consortium Contact: Auffray C
             CNRS UPR 420 - Genetique Moleculaire et Biologie du Developpement
             IFR 1221 - Rue Guy Moquet 19, Batiment G - BP 8
             94801 Villejuif Cedex, FRANCE
             Tel: +33-1-49 58 34 98
             Fax: +33-1-49 58 35 09
             e-mail: auffray@infobiogen.fr
             This clone is available royalty-free through IMAGE Consortium
             Distributors.
             IMPORTANT: This represents single read sequence of this IMAGE cDNA
             clone as supplied from the European distribution centers.
FEATURES     source
             1..562
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone_lib="Soares fetal liver spleen 1NPLS"
             /clone="IMAGE cDNA clone 123055, 5' end"
             /note="Matches EST Acc. No. R00125"
BASE COUNT   125 a 174 c 170 g 91 t 2 others
ORIGIN
alignment_scores:
  Quality: 503.50      Length: 162
  Ratio: 4.231         Gaps: 2
  Percent Similarity: 73.457 Percent Identity: 64.815
alignment_block:
US-09-068-377-1 x IROEST015 ..
Align seg 1/1 to: IROEST015 from: 1 to: 562
255 ThrLeuGluGlyCysAspValGluGlyAspIleAsnGlyPheIleGlnse 271
|||||
3  ACCTGGAGGCTGCAGTACAGCCGACATCGACAGTTCATCCAGGC 52
|||||
271 rlySerThrGlyArgGluProAlaProValProTyGlnAsnTyT 288
|||||
53 CAAGACACGGGCACAGACCCCGG..... 79
288 yzAspArgGluValThrProLeuIleGlySerProSerIleGlnProSer 304
79 ..... 79
305 CysGlyValIleIysArgPheSerGlyLeuLeuHisGlySerProLysTh 321
|||||
80 .....TTCTCTGGATTGCTGCACGGAAGTCCCAAGAC 111
|||||
321 rThrPro...SerAlaProAlaSerThrGluThrLeuThrProThrP 337
|||||
112 CACTTCGTTGGCAGCTTCTGCTGGCTCCACAGAGACCCCTGACCCCAACC 161

```

```

337 roGluArgAsnGluLeuValTyAlaSerIleGluValGlnAlaThrGln 353
|||||
162 CCAGCGGAATGAGGTGCTCATACAGCCATCGAGTGCAGGAGATACAG 211
|||||
354 GlyAsnLeuAsnSerAlaGlnAspTyArgAlaLeuTyArgPtyrTh 370
|||||
212 GGAACCCGGCCTCACAGCCAGGAGTACCGGGCGCTCTACGATTATAC 261
|||||
370 rAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspIleLeuAlav 387
|||||
262 AGCGAGAACCAGATGAGTGGACCTGTCACGGGAGACATCCTGGAGG 311
|||||
387 alIleLeuGluGlyGluAspGlyTyTrpThrValGluArgAsnGlyGln 403
|||||
312 TGATCTGGAAGGGAGGATGGCTGGTGGACTGTGGAGAGGACGGCAG 361
|||||
404 ArgGlyPheValProGlySerTyLeuGluLysLeu 415
|||||
362 CGTGGCTTCGCTCCTGGTCTTACCTGGAGAAAGCTT 397
|||||
seq_name: gb_p12:SPBC11C11
seq_documentation_block:
LOCUS      SPBC11C11  22500 bp      DNA      PLN      14-DEC-1999
DEFINITION S.pombe chromosome II cosmid c11c11.
ACCESSION  AL031528
VERSION    AL031528.1  GI:3560191
KEYWORDS   60s ribosomal protein l5; alp1; FCH domain; Fes/Cip4 homology;
             pseudouridine synthase; RNA binding; RNA recognition motif; rp15-2;
             SH3 domain; signal transduction; splicing protein; tRNA-pro;
             tRNA-Ser; tubulin-folding cofactor D homolog.
SOURCE     Schizosaccharomyces pombe
ORGANISM   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
             Schizosaccharomycetaceae; Schizosaccharomycetes.
REFERENCE  1 (bases 1 to 22500)
AUTHORS    Wood,V., Rajandream,M.A., Barrell,B.G., Beck,A. and Reinhardt,R.
TITLE      Direct Submission
JOURNAL    Submitted (03-SEP-1998) European Schizosaccharomycetes genome
             sequencing project, Sanger Centre, The Wellcome Trust Genome
             Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrellesanger.ac.uk
             and Max Planck Institut fuer Molekulare Genetik, Ihmestr. 73,
             D-14195 Berlin, Germany
COMMENT    Notes:
             Details of yeast sequencing at the Sanger Centre are available on
             the World Wide Web.
             (URL, http://www.sanger.ac.uk/projects/S_pombe/)
             During 1995 to 1996 about 66% of S. pombe chromosome 1 was
             sequenced by the Sanger Centre. The sequencing of the S. pombe
             genome is now being continued with funding from The European
             Commission. Fourteen European sequencing laboratories, including
             the Sanger Centre, are participating in the project.
             Protein coding regions (CDS) have been predicted with the help of
             computer analysis using the Genefinder program in Pombase (an ACEDB
             database) with additional predictions for the branch-acceptor sites
             supplied by the program Sp3splice. CAUTION: It is possible that for
             any individual CDS we may have underestimated or overestimated the
             number of introns/exons oremb1.tab we may not have chosen the
             correct splice donor/acceptor sites.
             CDS are numbered using the following system eg SPBC25H2.01c. SP (S.
             pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c
             (complementary strand).
             The more significant matches with motifs in the PROSITE database
             are also included but some of these may be fortuitous.
             The length in codons is given for each CDS.
             IMPORTANT: This sequence MAY NOT be the entire insert of the
             sequenced clone. It may be shorter because we only sequence
             overlapping sections once, or longer, because we arrange for a
             small overlap between neighbouring submissions.
             Cosmid c11c11 is overlapped by cosmids 17D1 and 388.
FEATURES     Location/Qualifiers
             1..22500
             /organism="Schizosaccharomycetes pombe"

```

```
/strain="972h-"
/db_xref="taxon:4896"
/chromosome="II"
/clone="cosmid cl1c11"
/map="IIR"
misc_feature
1..1769
/note="nominal overlap with cosmid SPBC17D1, EM:AL031322
S. pombe chromosome 2"
28..40
misc_feature
41..1496
/note="ctaattgtttag, splice branch and acceptor"
gene
41..1496
/label="SPBC11C11.01"
CDS
join(41..739,787..965,1016..1496)
/partial
/gene="SPBC11C11.01"
/note="SPBC11C11.01, len:452, SIMILARITY: LOW to Homo
sapiens, O15042, KIA0332, (1028 aa), fasta scores: opt:
176, E():0.00054, (22.1% identity in 249 aa)"
/codon_start=1
/label="SPBC11C11.01"
/product="hypothetical protein"
/protein_id="CAA20683.1"
/db_xref="GI:3560192"
/translation="SDDEDDIFLAGERKYLQNSYERKQKYVNSFNITKSVIRNE
HGNNLTQELLRHEAYERKEETANENDVDQGMFIQNNILNMLNLMKIFYHMDNIS
PYTNGSLKESLYKSYITAVNIFSTGDLKKGIAIFQDLSAQAVQYLSNCKIDGR
LISATITNPKRLPNAEHLSESTKTDESQDKDKLTKLDRAKLEWIOGLNCEKSGIG
NLLCAVEHVNHHVETDAFLKEFFNDOPDTDDTKVYDDYINERGEKKLSLYLMDNI
LFGLSGTSLVWRYRFSPEPHVRLDLDLYLSKRLGGRIKEDIFCKKVVKVIEYKWT
WIAQEEILERAWRFSGNTPOSPQINSALKVETKNSWTASETEGLQDDDEEINGI
PVDVNEELLNVEFISQPMSTETSSSSPOPTERKAKFKPSTQGTFTVSKRMHAE
L"
740..745
misc_feature
740..745
/gene="SPBC11C11.01"
/note="SPBC11C11.01, len:452, SIMILARITY: LOW to Homo
sapiens, O15042, KIA0332, (1028 aa), fasta scores: opt:
176, E():0.00054, (22.1% identity in 249 aa)"
/note="gtattg, splice donor sequence"
771..786
misc_feature
771..786
/gene="SPBC11C11.01"
/note="ctaattgtttag, splice branch and acceptor"
966..971
misc_feature
966..971
/gene="SPBC11C11.01"
/note="gtattg, splice donor sequence"
1001..1015
misc_feature
1001..1015
/gene="SPBC11C11.01"
/note="ctaattgtttag, splice branch and acceptor"
3189..5408
gene
3189..5408
/gene="SPBC11C11.02"
CDS
join(3189..3257,3352..3443,3509..5181,5230..5408)
/note="SPBC11C11.02, len:670,
SIMILARITY:Schizosaccharomyces pombe, CCI5_SCHPO, cell
division control protein 15, cell division control protein
15., (927 aa), fasta scores: opt: 647, E():1.2e-27, (26.1%
identity in 912 aa)"
/codon_start=1
/label="SPBC11C11.02"
/product="FCH and SH3 domain protein"
/protein_id="CAA20684.1"
/db_xref="GI:3560193"
/translation="MSQQLSFNASAKPKDSFSYFWGANDGYHALLSRSDVKHIN
ELRSGFYHERANIEDYAKRMALSKRTTSSLETGLKESVQVMKAEDVNMKASHLQI
SOLLQDDVNFATKLEAKLKVQSGIEKVKHDKLSKHQALYKAQDKYHLCCKY
NYVQSQNMFLKEELKNKLNQNAITASSDYQSAVAARVDSYARWNTMRWSTC
DKLQDIEERHRLSVVMTFTLLSRSCFNDQACERIKNLKQCSQDVLEPIDA
KSTGIGIPQPKFYDYKGEVDPDVSVELVQANFORAOTKLENDNPLNRPYVLSATAR
NESSFENLPNTPSQISLTIVSSNSNGSSPKKSFSLKFLSRESSTPNVGNATP
DALSPNDSPUTSAADQMKHLSQEEFKQNTPAAGFNSNLTLPNRYNELGSLP
SPNSVSDSPNTPSRNQIQIEFGVQLQMNRAVSPVYDSKRGSRSSFTLRK
SRSPKRPSSSQNSARLPSLTGCLNPNYDFGVDPDPAAGTATDDDEPTDRDSSF
VDDTNTKATGNTNRLSLPVPDGGDTSDINPTSDGORTILGVVSALYDYDAIPE
EISFRKGDITIAVLKLYEDGWEGFVVGEDDHNRGFPSPNFVREIEV"
3258..3263
misc_feature
3258..3263
/gene="SPBC11C11.02"
/note="SPBC11C11.02, len:1107"
/note="gtatgg, splice donor sequence"
3335..3351
misc_feature
3335..3351
/gene="SPBC11C11.02"
/note="ctaattgtttag, splice branch and acceptor"
join(3352..3443,3509..3665)
misc_feature
3444..3449
/note="Match to FCH, Score 94.19"
misc_feature
3444..3449
/gene="SPBC11C11.02"
/note="gtgaga, splice donor sequence"
3493..3508
misc_feature
3493..3508
/gene="SPBC11C11.02"
/note="ctaattcaacctag, splice branch and acceptor"
5182..5187
misc_feature
5182..5187
/gene="SPBC11C11.02"
/note="gtatgt, splice donor sequence"
5215..5229
misc_feature
5215..5229
/gene="SPBC11C11.02"
/note="ctaattgaattag, splice branch and acceptor"
5235..5399
misc_feature
5235..5399
/gene="SPBC11C11.02"
/note="Match to SH3, Score 59.53"
5802..5896
trna
5802..5896
/gene="trna-Ser, anticodon=GCT"
/note="aa score = 14, T-score = 8, length = 95, trna-Ser,
a nticodon=GCT"
5802..5896
gene
5802..5896
/gene="trna-Ser, anticodon=GCT"
6056..8022
gene
6056..8022
/gene="trna-Ser, anticodon=GCT"
join(6056..6083,6133..6316,6360..8022)
CDS
join(6056..6083,6133..6316,6360..8022)
/note="SPBC11C11.03"
/note="SPBC11C11.03, len:624, SIMILARITY:Homo sapiens,
Q14777, retinoblastoma-associated protein hcc., (642 aa),
fasta scores: opt: 409, E():2e-15, (22.7% identity in 585
aa)"
/codon_start=1
/label="SPBC11C11.03"
/product="coiled coil protein"
/protein_id="CAA20685.1"
/db_xref="GI:3560194"
/db_xref="SWISS-PROT:Q10198"
/translation="MQDSYVARYRYSQAPSSNLRITFTFGNGLGTSRTSLAPORTLV
NARQIDPGLSSRLVTPMRPSLAPNTRRSVRVSNASFISQTNITSIKPRPLSD
RYQOECATQVNVYLLSGESQIPLGNRMFSTREFAAIFKHLNKLDPNRFGARV
REDVTCLKALNYPFLDSISPSRLVAGSPHWPAILGLMLHVVSLIOCTEKAVMY
TVQNSLDDHLVDKVLFDYLVRYTHLYLDESPESEPEKELKATNQNDLYNTEA
LKSTNEELINQIKSAEELDSALQVLEERYTMQDEVKFQSAWSGKMSRSTNLMK
QLQVNIKESQQLLKEKRSKYQVENQDISISEFKMVSEKQDRNLNMIGKI
SELKRVFTDLLIQASIDSLSEKKVKQFNLSAYRIGIVPIAAIRSANNDFEINPEG
PNYINLDLKNKVRPFINEVRSITLFEHEQNKSLQEHVDTVNDLAELODELRG
ESRLTSVLSQNLRETASEKNAFDESKLRELQQLKLSHSHMLQDRIOSIN
TEADIAHACMEYKNNYKEVAFVLGEIHFHKLHVQDSLEDLKMDYQKLELDLSREL
6084..6089
misc_feature
6084..6089
/gene="SPBC11C11.03"
/note="gtatgt, splice donor sequence"
6117..6132
misc_feature
6117..6132
/gene="SPBC11C11.03"
/note="ttaacaataattag, splice branch and acceptor"
6317..6322
misc_feature
6317..6322
/gene="SPBC11C11.03"
/note="gtatga, splice donor sequence"
6343..6359
misc_feature
6343..6359
/gene="SPBC11C11.03"
/note="ctaattgtttag, splice branch and acceptor"
complement(8353..11933)
gene
complement(8353..11933)
/gene="SPBC11C11.04c"
CDS
complement(8353..11933)
9201..9583,9623..10609,10654..11742,11787..11933)
/gene="SPBC11C11.04c"
/note="SPBC11C11.04c, len:1107"
```

```

/codon_start=1
/label=alpi
/product="tubulin-folding cofactor d."
/protein_id="CAA20686.1"

alignment_scores:
  Quality: 302.50      Length: 639
  Ratio: 1.235        Gaps: 12
  Percent Similarity: 38.341  Percent Identity: 17.840

alignment_block:
US-09-068-377-1 x SPBCL1C11 ..

Align seg 1/1 to: SPBCL1C11 from: 1 to: 22500

46  ArgAlaGlnAlaGluArgTyrGlyLysGluLeuValGlnIleAlaAr 62
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3507 AGAGCGCAACATTCAAGAAGATTATGCCAGCGAATGCCAAGTTGCTCG 3556

62  glysAlaGlyGlyGlnThrGluMetAsnSerLeuArgThrSerPheAspS 79
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3557 AACACCTTCAGCTCTTTGGAACTGGTTGCCCAAGAAAGATCAGTCCAAG 3606

79  erLeuLysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAla 95
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3607 TTATGAAGAGCTGAAGTGTGACAATATGCGCAAGTCTCATTTGCCAGATTCT 3656

96  LeuAlaLeuArgGluGluLeu...ArgSerLeuGluGluPheArgGluAr 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3657 CAGTTGCTTCAGGATGATGTCGAAAATGGCTTACCGGTACGCTGCCCTC 3706

111 gGlnLysGluGlnArgLysLysTyrGluAlaIleMetAspArgValGlnL 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3707 TTTAAAGGACAAGAAAAGATGATGCTTTCCGGGATAGAAAAAGTTTCATA 3756

128 ySerLysLeuSerLeuTyrLysLysThrMetGluSerLysLysAlaTyr 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3757 AGGATAAATCTCCAAGCATCAAGCTCTTTGTTAAAGCTCAAGACAAATAC 3806

145 AspGlnLysCysArgAspAlaGluGlnAlaPheGluArgVa 161
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3807 CATTACTTTGCAAGAAAGTAAT.....TACTATGT 3838

161 lSerAlaAsnGly.....HisGlnLysGlnValGluLysSerGlnAsnL 176
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3839 TTCTCAGCAAAACATGCTTTTCGGAAGAGGCTFAGAGAACAATCGCA 3888

176 ySAlaLysGlnCysLysGluSerAlaThrGluAlaGluArgValTyrArg 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3889 AATTAACAAGACTCAGAATGCTATCAGCTAGTTCTTCTGCTACTATCAA 3938

193 GlnAsnIleGluGlnLeuGluArgAlaArgThrGluTrpGluGlnGluH 209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3939 TCAGTGTAGCAGCTGTGAGAGTTCCATACGCTCGTTGGACCAACGAATG 3988

209 sArgThrThrCysGluAlaPheGlnLeuGlnGluPheAspArgLeuThrI 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3989 GCGGTCAACCTGTGACAACTTCAGGACATTTGAAGAAGAAGAAGACATT 4038

226 leLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeuSerMetGlnCys 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4039 TCTCTGAAAAGCGTCATGTGGACATTTACACTACTTATCTCTCGTAGTTGC 4088

243 ValLysAspAspGluLeuTyrGluGluValArgLeuThrLeuGluGlyCy 259
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4089 TTTAATGATGATCAACATGTGAACGAATTCGAAAATCTTGAACAGTGC 4138

259 sAspValGluGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyA 276
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4139 CAGTGTCTCAGGATGTTTGGAAATTTATTGACGCAAAATCTACAGGAA 4188

276 rgGluProProAlaProValProTyrGlnAsnTyrTyrAspArgGluVal 292
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

4189	CGGAAATCCCTCAACACTCCAAAAATTTTATGACTATTATATAGGGCGAGGTT	4238
292	292
4239	CCCGATGATTCTGTAGAGTGTGGTTCAGGCTAATTTTCAGCGGGCCCAAC	4288
292	292
4289	CAAAATAGAAATGACAACATGCCATTGAATCGTCCATACGCTCTTCTG	4338
292	292
4339	CCACGCTCGTAATGAGTCGTCCTTCGAAATACCTTCCTAATACTCCT	4388
292	292
4389	AGTGCAATTCAATCGCTTACCACGTGTTCTTAACCTTCCCAAAATGG	4438
292	292
4439	TAGAAGCTCTCCAAAGAAAAGCTTTTGAGTAAGTTAAATTAACGAGTC	4488
293ThrProLeuIleGly.....	297
4489	GTCCCTCGACTCCAAACGCTGGTAATACGGCTCCGGATGCATTTATCAAGT	4538
298SerProSerIleGlnProSerCysGlyValIleLysAr	310
4539	CCGAGAAATGACTCTCCTTTAACTTCAGCGGCTGATGAACAGATGAACA	4588
310	gPheSerGlyLeuLeuHisGlySerProLys.....ThrThrProSerA	325
4589	TCTTTCT.....TTACAAGAAGACCCAAACAGAACTCTACGCCTGCTG	4632
325	laProAlaLaserThrGluThrLeuThrProThrProGluArgAsnGlu	341
4633	CTTCCTGGTGCCTTCCTTAACCTAATACACTACCACCAAGGATATAACGA	4682
342	Leu.....	342
4683	CTTGGTCTCTACCAAGTCCTAATTCGTGTTTCTTACCAGGAGCTCCG	4732
342	342
4733	CCCAATGTTAACACTCCATCAGCGGCTCAACAAATTCAGGAAGAATTG	4782
343ValTyrAlaSer	346
4783	GTAGTGTACTACAAATGGAGAATCGTGCGGTTTACCCGTTTATGATCA	4832
346	346
4833	AGAAAGATGGATCTCGAAGTTCATTTCATATTGAGGAAGTAGGTCAAC	4882
347IleGluValG	350
4883	AAAAGTCTCTCTTCCCTTCTCAGAATGCCAGCCGACTTCCTCGTT	4932
350	lnAlaThrGlnGlyAsnLeu.....	356
4933	CCTTAACACCAGAAATTTTAGAGCCAAATATGATTTGGTGTTCGAGTT	4982
356	356
4983	GATCCAGCAAGTGGTACTGCACCTACCGATCAGAGCCTTATACTGATCG	5032
356	356
5033	GGACTCTCTTGTGTGATGATACAATAAATACAAAGCTACTGGAACA	5082
356	356
5083	CGTCTAACCGACTTTCTTTTACGTGCATATCCTACAGACGGCGGTGACAC	5132

```

357 .....AsnSerSerAlaGlnAsp..... 362
|||||
5133 AGTATTGATAATCCAAACAAGTACAGATGGTCAACGATTCTTCTGTTATGG 5182
|||||
363 .....TyrA 364
|||||
5183 TATGCTTAATCTGTCATCGCTATTACAAATTACTAACTTGAATTAGTAT 5232
|||||
364 rGAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspLieser 380
|||||
5233 CTGCTTTATGATTATGATCTGCAATACCCGAAGAGATAAGCTTCAGG 5282
|||||
381 AlaGlyAspIleLeuAlaValIleLeuGluGlyGluAspGlyTyrTrrp... 396
|||||
5283 AAGGAGATACCAITGCAGTCTTAAAGCTTATGAAGACGCTGTTGGGA 5332
|||||
397 .....ThrValGluArAsnGlyGlnArgGlyPheValProGlys 410
|||||
5333 GGGATTTCTGCTGGAGAGATGACCATACCGTGGCAATTTCCGTCAA 5382
|||||
410 erTyrLeuGluLysLeu 415
|||||
5383 ATTGTGTCGTGAATC 5399
|||||
seq_name: gb_pr4:AF128536
seq documentation_block:
LOCUS AF128536 3255 bp mRNA PRI 26-JUL-1999
DEFINITION Homo sapiens cytoplasmic phosphoprotein PACSIN2 mRNA, complete cds.
ACCESSION AF128536
VERSION AF128536.1 GI:5305705
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 3255)
Ritter,B., Modregger,J., Paulsson,M. and Plomann,M.
PACSIN 2, a novel member of the PACSIN family of cytoplasmic
adapter proteins
JOURNAL FEBS Lett. 454 (3), 356-362 (1999)
MEDLINE 9938785
REFERENCE
2 (bases 1 to 3255)
Ritter,B., Modregger,J. and Plomann,M.
Direct Submission
JOURNAL Submitted (15-FEB-1999) Institute for Biochemistry II, University
of Cologne, Joseph-Stelzmann-Str. 52, Cologne D-50931, Germany
FEATURES
source
1..3255
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="adult retina; infant brain"
209..1669
/note="related to PACSIN1; contains CDC15 N-terminal
domain, 3 EH domain-binding NPF motifs, and a C-terminal
SH3 domain"
/codon_start=1
/product="cytoplasmic phosphoprotein PACSIN2"
/protein_id="A041781.1"
/db_xref="GI:5305706"
/translation="MSVTVDDSDVGVESDSEWYGNKRTVKRIDDGHRLGSLMNC
LHERIKAYAQQLTEWARWRQLVERGPQYGTVEKAWMAPMSAEKRVSELHLEVKA
SLMNDDEKIKNMQEAFPHKMMGGFKETKEADEGRKAOKPWAKLKEVEAAKKAHH
AAKREKLAISREANSKADPSNPQLKLODKIECKQDVLTKKEKESLKELDQG
TPQYENNEQVEQCCQFEERKLFREVLVQLHNLNSVAGTKAHLHLEQSIRA
ADAPEDLRFNRHNGPWAMNPQSEWSADLIRLSRREKKATYDGTFTLLTGINGTD
QFLSKPSSTLVNPSNPAQSQSSYNPFDEDDTSTVSEKEDIKAKNTSSYEKTQ
SYPTQSDDESNNPSSFDANGDSNPFDDDTSGTEVRVRALYDVEGOEHDLSFKAG
DELTMEDEQGWCKGRLDNGQVGLYPANYVEAIQ"
polyA_signal 794 a 792 c 934 g 735 t
BASE COUNT
ORIGIN

```

```

alignment_scores:
Quality: 301.00 Length: 493
Ratio: 1.103 Gaps: 20
Percent Similarity: 55.375 Percent Identity: 23.529
alignment_block:
US-09-068-377-1 x AF128536
Align seg 1/1 to: AF128536 from: 1 to: 3255
9 AspAlaPheTrpCysArgAspPheThrAlaHisThrGly...TyrGluVa 24
|||||
254 GACAGCTTCTGG.....GAGTCGGGAACACTACAAGCG 285
24 lleuLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValGluG 41
|||||
286 GACTGTGAAGCGGATCGACGATGCCACCGCTGTGCAGCGACCTCATGA 335
41 luLeuLeuArgGlnArgAlaGlnAlaGluGluArgTyrGlyLysGluLeu 57
|||||
336 ACTGCTTCATGACGGCGCGCATCGAAGGCGTATGCCGACGACGCTC 385
58 ValGlnIleAlaArgLys.....AlaGlyGlyGlnTh 68
|||||
386 ACTGAGTGGCCCGCGCTGGAGCGACGCTTCTGGAGAAAGGCCCCAG... 433
68 rGluMetAsnSerLeuArgThrSerPheAspSerLeuLysGlnGlnThrG 85
|||||
434 ...TACGGGACCGCTGGAGAAGGCGCTGATGCTTCCGAGGCG 479
85 luAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuGluGlu 101
|||||
480 AGAGGTGAGGAGCTGCACCTCGAGGTGAAGGCTCCTCCTGATGAACAT 529
102 LeuArgSerLeuGluGluPheArgGluArgGlnLysGlu..... 114
|||||
530 .....GACTTCGAGAAGATCAAGAACTGGCAGAAGGAGGCTTTTCAAA 573
115 .....GlnArgLysLysTyrGluAlaIleMetA 124
|||||
574 GCAGATGATGGCGGCTTCAAGGAGACCAAGAAAGCTGAGGAGCGCTTC 623
124 spArgValGlnLysSerLysLeuSerLeuTyrLysLysThrMetGluSer 140
|||||
624 GGAAGGCACAGAGCCCTGGCCCAAGAGCTGAAAGAGGTAGAAGCAGCA 673
141 LysLysAlaTyrAspGlnLysCysArgAspAlaAspAla.....G 155
|||||
674 AAGAAAGCCACCATGCAGCGTGCAGGAGGAGGAGGCTGCTATCTCACG 723
155 uGlnAlaPheGluArgValSerAlaAsnGlyHisGlnLysGlnValGlu 172
|||||
724 AGAAGCCAAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 773
172 ysSerGlnAsnLysAlaLysGlnCysLysLysLysLysLysLysLys 188
|||||
774 AATTGCAAGACAAAATAAGAAAGTCAAGCAAGAGTCTTAAAGCCAAA 823
189 ArgValTyrArgGlnAsnIleGluGlnLeuGluArgAlaArgThrGluTr 205
|||||
824 GAGAAGTATCAGAAGTCCCTCAAGGAACCTCGACCGGCGCACACCCAGTA 873
205 pGluGlnGluHisArgThrThrCysGluAlaPheGlnLeuGlnGluPheA 222
|||||
874 CATGGAGAACATGAGCAGGTGTGAGCAGTCCACGAGCTTCGAGGAGA 923
222 spArgThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeu 238
|||||
924 AACGCTTCCTCTCCGGGAGGTTCTGCTGGAGGTTTCAAGACACCTA 973
239 SerMetGlnCysValLysAsp...AspGluLeuTyrGluGluValArgLe 254

```


974 AACCTGTCATATGGCTGGTGTACAAAGCCATTACCATGACCTGGAGCA 1023
254 uthrLeuGluGlyCysAspValGluGlyAspIleAsnGlyPheIleGlns 271
1024 GAGCATCAGCAGCTGATGACGAGTGGAGACCTGAGTGGTCCGAGCCA 1073
271 er_LysSer.....ThrGlyArg..... 276
1074 ATCAGCGGCCAGGCATGGCCATGAACCTGCCGAGTTTGAGGAGTGCTCC 1123
277GluProProAlaProValProTyrGlnAsnTyrTyrA 289
1124 GCAGACCTGATTCGACACCTCAGCCCGGA.....G 1152
289 sPArgGluValThrProIleGlySerPro..... 299
1153 AGAAGAGAGAGAGCCACTGACGGCTTACCTGACGGGCATCAACCAGA 1202
300SerIleGlnProSerCysGlyValIleLysAr 310
1203 CAGGGACACAGTTTTCGCCGAGTAAAGCCAGCAGCACCTTAATGTCCG 1252
310 qPheSer.....GlyLeuLeuHisGlySerProLysThrThrProSer. 324
1253 AGCAACCCGCCCGCTGTCGGAGTCACAGTCCAGCTACAAACCCCTTCGA 1302
325AlaProAlaAlaSerThrGluThrLeu... 333
1303 GGNTGAGGAGCAGACCGGGCAGCAGCTCAGTGAGGAGGAGCATTAAGG 1352
333 333
1353 CCAAAAATGTGAGCAGCTACGAGACAGCCAGAGCTATCCACCGACTGG 1402
334ThrProThrProGluArgAsnGluLeu.Val 343
1403 TCAGAGGATGATCTACAAACCCCTTCTCTCCACGGATGCCAATGGGA 1452
344 TyrAlaSerIle.GluValGlnAlaThrGlnGlyAsnLeuAsnSerSera 360
1453 CTCGAATCATTCGAGGAGCAGCCACCTCGGG.....ACGG 1490
360 laGlnAspTyrArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGlu 376
1491 AAGTGCAGTCCGCCCTGTATGACTATGAGGGCAGGAGCATGATGAG 1540
377 LeuAspIleSerAlaCysAspIleLeuAlaValIleLeuGluGlyGluAs 393
1541 CTGAGCTTCAAGGCTGGGGATGAGTGCACCAAGATGGAGGAGGATGA 1590
393 pGlyTrpTrpThrValGluArg.....AsnGlyGlnArgGlyPheValp 408
1591 GCAGGGCTGTCGACAGGAGCGCTGGACACACGGCAAGTTGCCCTATACC 1640
408 roGlySerTyrLeuGluLysLeu 415
1641 CGGCAAAATTATGGAGGCGATC 1663
seq_name: gb_pl1:SPDCD15

seq_documentation_block:
LOCUS SPDCD15 3993 bp DNA PLN 22-JUL-1997
DEFINITION S.pombe cdc15 gene.
ACCESSION X86179
VERSION X86179.1 GI:773262
KEYWORDS CDC15 gene; phosphoprotein.
SOURCE fission yeast.
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Archiascomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces.
REFERENCE 1 (bases 1 to 3993)

AUTHORS Fankhauser,C., Reymond,A., Cerutti,L., Utzig,S., Hofmann,K. and Simanis,V.
TITLE The S. pombe cdc15 gene is a key element in the reorganization of F-actin at mitosis
JOURNAL Cell 82 (3), 435-444 (1995)
MEDLINE 95360987
REMARK Erratum: [[published erratum appears in Cell 1997 Jun 27;89(7):1185]]
REFERENCE 2 (bases 1 to 3993)
AUTHORS Simanis,V.
TITLE Direct Submission
JOURNAL Submitted (10-APR-1995) V. Simanis, ISREC, Chemin des Boveresses, 155, CH-1066 Epalinges, SWITZERLAND
FEATURES
source
1..3993
/organism="Schizosaccharomyces pombe"
/variety="wild type"
/strain="972 h minus"
/db_xref="taxon:4896"
/chromosome="1"
376..3465
/gene="cdc15"
376..456
/gene="cdc15"
/number=1
join(376..456,515..606,799..3236,3293..3465)
/gene="cdc15"
/function="colocalises with f actin at mitosis but not in interphase"
/codon_start=1
/product="phosphoprotein"
/protein_id="CAA60115.1"
/db_xref="GI:2275095"
/db_xref="SPTREMBL:O14365"
/translation="MEYNGVQSSEAAPVYTKSSVKFRDNFNGSDAGMDALMSPTKSS LSVLESDIEFYAKRASIERYASKLOELAAASADIPEVGSILNILLMSERKDAYEYS HEEVYQQINTELRNRIEYIDTQEQKVAAANAIEELYOKKTALEIDLITSEKRDAYS CNKINSYMRQTKTGRELKYNLIRQAALAVKMDAEYRETNELLITVTREWIDRW TEVCDFAQHIEEYLEFLKTNWAYANIISTACVKDDSECKIRLTENTIDITQ MIONEGTCTTIPPLPEFNDYFKENGLAYDIDQLISKAPSYPYSSRRPSASLASPT RSAPRKTSSTVSSEVSSPPTSLHSPVKPVSNQEQVQTEVELSIPTVQEAESQ KPVLTGSSMRPSTVPTFEVAARPLTSMVRSNAETVQAIPTAATDIDPEVKEGK NSENAITDNDIDILSLOLQPTATGSSRSRLFSRHGHGQSITSLGSIKRSKIMERMGR PTFPMGSSFNMSRSTPTKEGFASQNHQATGASVQSGDELEDIDPRANVVLNVPNM LSVGEAPVSTSKEDKVDPPIANAMAEISSMRROSTVSDDEAPVSLKTSSTR LINGIYHSRNTSIAISIDGVKPKSTLGAAPAAHTSAQMORMNSFASOTKOVFGPORT ENSARESLRHSRNSMRSPMLSRSTLRPSERSASSLSVRQSDVVSFAPSTRAR GQSVGGQRPSSMSLYEYKNSQPLSMQKSVSPNLPGRRRSSSVLOSOKRSSNT SNRNGYSGSRPSSEMGHRYGSGSRMQSVQRSTSRARSPEPTNRNSVOSKNVDP RATFTAEGEPILGYVIALDYQAQIPEEISFQKGTLMVLRLTQEDGWDGELINVPNS KRGLFPSNFTVTV"
515..606
/gene="cdc15"
/number=2
799..3236
/gene="cdc15"
/number=3
846..907
/gene="cdc15"
/note="SH3 domain"
3293..3465
/gene="cdc15"
/number=4
BASE COUNT 1239 a 781 c 781 g 1192 t
ORIGIN
alignment_scores:
Quality: 297.00 Length: 326
Ratio: 1.547 Gaps: 7
Percent Similarity: 58.896 Percent Identity: 26.380
alignment_block:

3.

/product="dna ligase"
/protein_id="CAB08595.1"
/db_xref="GI:2094856"
/db_xref="SWISS-PROT:P12000"
/translation="DQAAFCEVKYDGERAQVHTEGKVFVESRNSENVRYPDLS
VYSKWKPPARSEFIDCEAVGWRDNDENKILPFQOKLATRRKQVKGIDIKVRACLFAP
DILYLNQPLEETPLNKRLLLYSMFQSTGDFTFARHSDQKSIIEEFLEESVKDS
CEGLVMKMLGPOSHYEPKRSRHLWKVDYLSGVGDSLDLIVIGAYVKGKRTSVY
GALLGCDPDPTETVQICIKLGTGTFSEEHLETYNOLKDIVISKKDFYAHSDVPAHQ
PDWFEKYLWEVLAADLSLSPVYKAAIGYVQEDKGISLRFPFIRIREDKSWEDAIT
SEOVSEFYRSQVAYSQKEGKESPAEDY"
1. 1089
/gene="cdc17"
/note="PS00697 ATP-dependent DNA ligase AMP binding site"
484. .567
/gene="cdc17"
/note="PS00333 ATP-dependent DNA ligase signature 2"
1374. 3647
/gene="SPAC20G8.02"
1374. .3647
/gene="SPAC20G8.02"
/note="SPAC20G8.02, len:757, SIMILARITY:Saccharomyces
cerevisiae, Q12204, chromosome xv reading frame Orf
yoc022c., (715 aa), fasta scores: opt: 370, E():1.2e-15,
(29.6% identity in 766 aa)"
/codon_start=1
/label="SPAC20G8.02
/product="hypothetical protein"
/protein_id="CAB08596.1"
/db_xref="GI:2094857"
/db_xref="SPRMBEL:P87109"
/translation="MILYIVLPFYVTRTKPYSLPSSLOWITNMCWLSPIYILCHC
SKHOETIILAKPLIAMNSEEPLKRVOPPKLSFRHLISRSVTDAPSILRVWFYAV
DRLRKSTETKAKNPLPFSAESEHIEKSLKAVENDGSEPVNVNEDLYYSV
NVYSRELSPYWDQPVVILRGTFWFSFGDKLXPCEENLATOVPEGVNSCPREFSN
EKDSAAQKTMALLGRYTGTFVQYTGSRNARLVYDDFYRNVSVKIMNRESFASFHRS
DKLYRGYELMDLENSRSPVPTFEILTSTLLNDSDPSDNPSTESTIDLPSAT
DASHLMRDESNVHLLICCHGIGQKMGERVEVTSFYVKDISNFRKTLKTFNNSPDLQ
AVYPLKGGNGVQCLELLWRQDIRFGMARLDSSFADDDDDDESINMRDLALDDL
EDDSIPTLDINIPITGILRNIISDVLVDLLVQCPNVRDKILAAVVRKLNRLNLYK
KNVPSFNGHSLGSLGALILFDIIRYQGNIKYSKLQDPFPVANFALGSLGLFQM
LNGKKIAGPIKTNLTBSLSYSEQSDSGVSLSCQNFALIFHTDPLSYRVEPLVYK
QMARLKPKISHPHODLSSGVGHKIAGALNVLSGLSGIANTLILKLSLTSASF
NEATADSHDSQGAENIRDWHIDERMRYLRNKTGRIDPMLQEGALDTSYSYVAMNAHS
EYKNVDLAHFILTQL"
4633. .6306
/gene="itr2"
4633. .6306
/note="SPAC20G8.03, len:557"
/codon_start=1
/label="itr2"
/product="myo-inositol transporter 2"
/protein_id="CAB08597.1"
/db_xref="GI:2094858"
/db_xref="SWISS-PROT:P87110"
/translation="MDFNNIPLATIKSSDKGKDFIVEMTTRPSETKKKVPFSDMRE
IPSLPNEEANATPOANEVADENGFEAEKISSWJWLVSAGVIGSLFGYDTGVI
SGALVLDGLHVLSSCKOKELITSATSFALISSTWISHLADVGRKRLILCCALDAIF
VIGSVMAASRNVMVYGRFIVIGIGLTSLSIVPMITITELAPLRKRLVLIYVFI
TGGOLIAYSINAAEPVHOGWRMFGIAGAAPALGOLISLEWTPESPRIYLRHNNKRV
YKLSLRHPKAPAEIAYKVSLLIQEGVDFPEGNKQHFHSLKVLFTVPSNRNSLF
IGCFLOWFOQSGTNAIQYFSAILIQSVGFKNSISVIVGATNFVTIVAFWFIIDRI
GRRILICTSAVIMAGLACAIAVHELPAITNTNSGWOYVVLASIIIFLASVASGI
GNIPWQQAELFPNEVRALGAFSTAINWGNLILISASFLTWMSITPTGTFALPAGFC
FVGLVTSYFTIPLAGMSIENIHKLLEKGFQWAVKESTKRVKRGRIDEA"
4882. .6234

misc_feature
/gene="itr2"
/note="Match to PF00083 sugar_tr, Sugar (and other)
transporter Score 471.91"
5050. .5100
/gene="itr2"
/note="PS00216 Sugar transport proteins signature 1"
5785. .5838
/gene="itr2"
/note="PS00216 Sugar transport proteins signature 1"
complement(6762. .8660)
/gene="SPAC20G8.04c"
complement(6762. .8660)
/gene="SPAC20G8.04c"
/note="SPAC20G8.04c, len:632, SIMILARITY:Homo sapiens,
ETFD_HUMAN, electron transfer flavoprotein-ubiquinone
oxidoreductase precursor(ec 1.5.5.1), (617 aa), fasta
scores: opt: 2054, E():0, (54.1% identity in 556 aa)"
/codon_start=1
/label="SPAC20G8.04c
/product="probable electron transfer
flavoprotein-ubiquinone oxidoreductase precursor(ec
1.5.5.1)"
/protein_id="CAB08598.1"
/db_xref="GI:2094859"
/db_xref="SWISS-PROT:P87111"
/translation="MSYLSRSALARSVGAKHLTGVLIRKIGRRGSRMHVPLASPSTL
LKISSQTLRQDFTVLGARFNHSSSVRLNLTDLNLRKLDTIEREVEDVDVICVAGPAG
LSAIRIKQQAANKARDIRVVVLEKAAEPGNHVSAGVIOPTALDELLPNRDDPPEN
CTAVTHDKMKEFLIKLHPPIPVPPAMKNHGNVMSLAETFWLKAFAEYGVETIYPSF
AASEVLNKGDSVIGVATINDFGVDSKGLPKDNFERGMFAHPVTLFASGAGHSLSKSI
IKRNLRGNCPEQTYGLGVKEVRVPDNEFKVEAHTLGPWRMDTYGGGFMYQFGD
NVTYVGLVLDYPNPYSPALPEQRMKQNFPAKVLKGGKCLFYAARALNEGQYQAI
PKLVPGGALIGCSAGFVNVAKIKGTHTKMSGIVAAIDAFDGRDAASKPLLLNDY
EENLKNYVFKELYSVRNI RPSFHSFLNGYGMAYSAVEAYVLGKRVPTLKHGGDA
KATKSASKYKPINYPKPDNVLSEDIPTPSYRSATPMHAENQPCHLFDHPRKDKSCFET
YKGVNFCPPAGTYEYVNDSEASSYKGRFVINSQVCHCKTCDIKDIPLOQWKTPOGG
DGPXYTLT"
9337. .9673
/note="TF1"
complement(10535. .13623)
/gene="cdc15"
complement(join(10535. .10707,10764. .13201,13393. .13484,
13543. .13623))
/gene="cdc15"
/note="SPAC20G8.05c, len:927"
/codon_start=1
/label="cdc15"
/product="phosphoprotein."
/protein_id="CAB08599.2"
/db_xref="GI:4726010"
/db_xref="SWISS-PROT:Q09822"
/translation="MEVNGVSOSEAAPYVTKSSVKFRDNFVSGSDAGMDALMSRTKSS
LSVLEIDFVAKRASIETERYASKLOELAASSADIPEVGSITLNNISMRITSGMAKA
HEEVSQINTELNRKIREYIDQEQQVVAANAIEELYQKXTALEIDISEKKDAEYS
CNKLSYMRQTKMTRKRELKYNLKIQAALAVKKMDAEYRETNELLTIVTREDWRD
TEYCDAFQHIEEYERLEFKTNMAYANIISTACVKDDECEKIRLTLENTIDEDITQ
MIQNEGCTTIPPLPEFNDEYKENGNDYIDQLISKADSPYSSRPSASASLASSPT
RSAPRKSTETVSVSEVSSPPTSLHPSPVKPVNEOVQVTEVELSIPVSIQAEASO
KPVLTGSSMRPVSPTSPTFEVAARPLTSMVRSRSHNAETEVOQAPAAATDIPVEKECK
NSENAITKNDDDIILSSQLQPTATGSRSSLSFNRHGHGSGTSLGSIARKKSIEMRWGR
PTSPFMGSSSTSNMGSRSPTTKEGFASQNHATGASVQSELEIDIPRANVYLVGNPM
LSVGEAEPVESTKEEDKVDPPDIANAAMELSSSMRRQSTSDVDEAPVYSLKTSSTSR
LNGVYHSRNTSIASIDIGVPKFTLGAPPAHTSAOMORMNSFASOTKOVFCGEORT
ENSARESLRHSRNSRSPSMLSRSTLRPSERSASSLSVQSDVVSFAPSTRAR
GQSVSGQQRSSMSLYGENKQSPQLSMQKRSVPNPLGNRRSSSVLQSQKSSSTNT
SNRNGYTSRPSRSGHRYGSGMRQVSRQSRSTRARSPEPTNRNSVQSKNDVP
RATPTAGEILGVIALDYQAQIPEISFQKGTMLVLRTQEDGWMDEIINVPNS
KRGLFPSNFMQTV"

alignment_scores:
Quality: 297.00 Length: 326
Ratio: 1.547 Gaps: 7
Percent Similarity: 58.896 Percent Identity: 26.380

12415 TTTCGGGTCCCTCAGCTTCTGCATCTTTGGCAAGTTCGCCAACACGTAGCG 12366

alignment scores:

alignment scores:

Quality: 284.50 Length: 493
 Ratio: 1.062 Gaps: 16
 Percent similarity: 54.361 Percent identity: 21.907

alignment_block:

US-09-068-377-1 x AF139492 ..

Align seg 1/1 to: AF139492 from: 1 to: 2079

```

 9 AspaAlaPheTrpCysArgAspPheThrAlaHisThrGly...TyrGluVa 24
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 GACACCTTCTGG.....GAGTCCGGGAAGTACAAACG 216

24 lleuLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValGluG 41
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217 GACTGTGAAGCGGATGTGATGGCCACCGCTGTGTGTGACCTCATGA 266

41 luLeuLeuArgGlnArgAlaGlnAlaGluGluArgTyrGlyLysGluLeu 57
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
267 ACTGTTTGCAGGAGCGGCATATCGAGAAGGCATATGCACAGCAGCTC 316

58 ValGlnIleAlaArgLys.....AlaGlyGlyGlnTh 68
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
317 ACTGAGTGGCCCGACGCTGGAGACAGCTTGTAGAGAAGGGACACAG... 364

68 rGluMetAsnSerLeuArgThrSerPheAspSerLeuLysGlnGlnThrG 85
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
365 ...TATGGGACCGTGGAGAAGCGCTGGATGGCTGTCATGCTGAGGCAG 410

85 luAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGluGlu 101
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
411 AGAGAGTGAGCGGAGTACACCTGGAGGTGAGGCATCATTGATGAACGAA 460

102 LeuArgSerLeuGluGluPheArgGlnLysGlu..... 114
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
461 .....GACTTTGAGAAGTCAAGAAGTGGCAGAGGAGGCTTTTCACAA 504

115 .....GlnArgLysLysTyrGluAlaIleMeta 124
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
505 GCAGATGATGGGAGCTTCAAGGAGACCAAGGAAGCAGAGGATGGCTTTC 554

124 spArgValGlnLysSerLysLeuSerLeuTyrLysLysThrMetGluSer 140
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
555 GGAAGCCGAGAGCCCTGGCCGAGAGAGCTGAAAGAGGTGGATGGCGCA 604

141 LysLysAlaTyrAspGlnLysCysArgAspAlaAspAla.....GI 155
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
605 AAGAAGCACACACACACAGCTGCAAGAGGAGAGAGCTGGCGCTCCTCCG 654

155 uGlnAlaPheGluArgValSerAlaAsnGlyHisGlnLysGlnValGluL 172
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
655 GGAAGCCCAACCAAGCAGAGATCCATCCCTCAACCTGAGCAGCTGAAGA 704

172 ySerGlnAsnLysAlaLysGlnCysLysGluSerAlaThrGluAlaGlu 188
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
705 RAATTCAGACAGATAGAAAGTGCAGAGGAGCTTCTAAAGACCAAG 754

189 ArgValTyrArgGlnAsnIleGluGlnLeuGluArgAlaArgThrGluTr 205
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
755 GACAAGTATGAGAGGCCCTGAGAGAGCTTGACCAGACTACACCCAGTA 804

205 pGluGlnHisArgThrThrCysGluAlaPheGlnLeuGlnGluPheA 222
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
805 CATGGAGAACAATGAGCAGGTCTTCAGCAGCTGCCAGCAGTTTGAAGA 854

222 spArgLeuThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeu 238
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
855 AGCGCTCGCTGCTTCCGAGAGAGTTCTGCTGGAGGTTTCAGAAGCAGT 904

239 SerMetGlnCysValLysAspAspGlu...LeuTyrGluGluValArgLe 254
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
905 GACCTGTCCATGTGGCTAGCTACAAAGGCATTTACCGGGAGCTGGAGCA 954

```

```

254 uThrLeuGluGlyCysAspValGluGlyAspIleAsnGlyPheIleGlnS 271
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
955 GAGCATCAAAAGCAGCTGATGCTGTGGAGCACCTGAGTGGTTC..... 997

271 erLysSerThrGlyArgGluProAlaProValProTyrGlnAsnTyr 287
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
998 .....CGAGCTAACCATGGCCAGCATGGTATGAACCTGGCCACAGTTT 1042

288 TyrAspArgGlu..... 291
  |||:|||||:

1043 GAGGATGAAGAGTGGTCTCCAGATCTGAATCGAACTCTCAGCCGAAGAGA 1092

292 .....ValThrProLeuIleGlySerProSerI 301
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1093 GAAGAAGAAGGCTGCTGATGGTGTCAACC...CTAACAGGAATCAACAGA 1139

301 leGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGly 317
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1140 CAGGTGACCAGTCTGGACAGACAAGCCCTAGCAGCAACCTT..... 1180

318 SerProLysThrThrProSerAlaProAlaAlaSerThr..... 330
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1181 .....AGTGTCCGAGCAACCCGCCAGTCCACGCACTACAGTC 1221

331 .....GluThrLeuThrProThrProG 338
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1222 CAGCTACAACCCCTTCGAGGAGCAGCAGCAGCGGCGCAGCGTCAGTG 1271

338 luArgAsnGluLeuValTyrAlaSerIleGlu..... 348
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1272 AGAAGGAGGACATTAAAGCCAAAATGTGACGAGCTATGAGAAGACTCAG 1321

349 .....ValGlnAl 351
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1322 AATTACCCCGCTGACTGGTCTGATGATGAGTCTAACAAACCTTCTCCTC 1371

351 aThrGlnGlyAsnLeuAsnSerSerAlaGlnAsp..... 362
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1372 CACGGATGCCAATGGGGAGCTCTAACCCGTTTGACGAGGACACTACCTCAG 1421

363 .....TyrArgAlaLeuTyrAspTyrThrAlaGlnAsnSer 374
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1422 GAACAGAACTGGAGTTCGAGCCCTCTATGACTATGAGGCAAGAACAC 1471

375 AspGluLeuAspIleSerAlaGlyAspIleLeuAlaValIleLeuGlu 391
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1472 GATGAGCTGAGCTTCAAGCTGGGGATGAGCTGACCAAGATAGAGGATGA 1521

391 yGluAspGlyTyrTrpThrValGluArg.....AsnGlyGlnArgGlyP 406
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1522 AGATGAACAGGGTTGGTGCAGGAGGCTTTAGACAGTGGCCAGTTTGCC 1571

406 heValProGlySerTyrLeuGluLysLeu 415
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1572 TATACCCAGCAACTATGTGAGGCTATC 1600

```

seq_name: gb_to:AF128535

seq_documentation_block:

LOCUS AF128535 3217 bp mRNA ROD 26-JUL-1999

DEFINITION Mus musculus cytoplasmic phosphoprotein PACSIN2 mRNA, complete cds.

ACCESSION AF128535

VERSION AF128535.1 GI:5305703

KEYWORDS house mouse.

SOURCE

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE

1 (bases 1 to 3217)

Ritter, B., Modregger, J., Paulsson, M. and Plomann, M.

PACSIN 2, a novel member of the PACSIN family of cytoplasmic

adaptor proteins

JOURNAL FEBS Lett. 454 (3), 356-362 (1999)

```

MEDLINE 99358785
REFERENCE 2 (bases 1 to 3217)
AUTHORS Ritter,B., Modregger,J. and Plomann,M.
TITLE Submitted (15-FEB-1999) Institute for Biochemistry II, University
JOURNAL of Cologne, Joseph-Steinmann-Str. 52, Cologne D-50931, Germany
FEATURES
source
1. .3217
/organism="Mus musculus"
/strain="C57BL6 x DBA"
/db_xref="taxon:10090"
251. .1711
/notes="related to PACSIN1; contains CDC15 N-terminal
domain, 3 EH domain-binding NPF motifs, and a C-terminal
SH3 domain"
/codon_start=1
/product="cytoplasmic phosphoprotein PACSIN2"
/protein_id="A041780.1"
/db_xref="GI:5305704"
/translation="MSVTYDSDVSGVEVSSDFWYGNKRYKTRIDGHRGCDLMNC
LHERARIEKAYAQQLTEWARRRWQLVEKGPQYGTVEKAWIAVNSEAERYSELRLEVKA
SLMNEDEFKIKWKEAPKHKMGKGFKEADGFRKAKPKWAKLKEVEAAKKAHH
TACKKEKLAISREANSKADPSLNPKLKLQDKIEPKODVLTKDKVEKSKELDOT
TQYVEMNEQVFEQCOQFEKRLREFREVLVQKHLDLSNVASIKYIYRELEQSLKA
ADAVEDLRFNRAHFGPMANWPFQEWESADLNRLSRKKKAVDGVTLTGINOTGD
QSQNKPGSNLSPVNPAGSTQLQSYNPFDEDDTGSISEKEDIKAKNVSSYEKTO
TYPTDMSDDSNPFSSTDANGSNPFDFTTSGTEVRVRLALDYEGOBHDELSFKAG
DELTKIEDDEQGWCKGRLDGQVGLYPANYVEAIQ"
polvA_signal 3179. .3184
BASE COUNT 845 a 771 c 863 g 738 t
ORIGIN

alignment_scores:
Quality: 283.50 Length: 494
Ratio: 1.086 Gaps: 15
Percent Similarity: 52.834 Percent Identity: 21.660

alignment_block:
US-09-068-377-1 x AF128535 ..
Align seg 1/1 to: AF128535 from: 1 to: 3217
9 AspaPheTrpCysArgAspPheThrAlaHisThrGly...TyrGluVa 24
|||||
296 GACAGCTTCTGG.....GAGGTGGGAAGTACAAAGC 327

24 lLeuLeuGluArgLeuLeuAspGlyArgLysMetCysLysAspValGluG 41
|||||
328 GACTGTGAAGCGGATTGACGATGCCACCCCTGTGTGGTGACCTCATGA 377

41 luLeuLeuArgGlnArgAlaGlnAlaGluGluArgTyrGlyLysGluLeu 57
|||||
378 ACTCTCTGATGACGGCGGCGGATCGAGAGCGGTATGCACAGCAGCTC 427

58 ValGlnIleAlaArgLys.....AlaGlyGlyGlnTh 68
|||||
428 ACTGAGTGGGCCCGCCAGCCTGGAGCAGCTGGTAGAGAAGGACACAG.. 475

68 rGluMetAsnSerLeuArgThrSerPheAspSerLeuLysGlnThrG 85
|||||
476 ...TATGGACCGTGGAGAAGGCTGGATAGCTGTCATGCTGAAGCAG 521

85 luAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGluGlu 101
|||||
522 AGAGGGTGAGTGAAGTGCACCTGGAGTGAAGGCATCACTGATGAATGAA 571

102 LeuArgSerLeuGluGluPheArgGluArgGlnLysGlu..... 114
|||||
572 .....GACTTTGAGAAGATCAGAACTGGCAGAAAGGACCTTTCACAA 615

115 .....GlnArgLysLysTyrGluAlaIleMetA 124
|||||

```

```

616 GCAGATCATGGAGCGCTTCAAGAGAGACCAAGAGAGAGATGGCTTTC 665
124 spArgValGlnLysSerLysLeuSerLeuTyrLysLysThrMetGluSer 140
|||||
666 GCAAGGCCAGACCCCTGGCCCAAGAGCTGAAGAGGTTGGAAGCGGCA 715
141 LysLysAlaTyrAspGlnLysCysArgAspAlaAspAspAla.....G 155
|||||
716 AAGAAGCGCCACACACAGCGTCAAGAGAGAGAGAGAGTGGCCATCCGC 765
155 uGlnAlaPheGluArgValSerAlaAsnGlyHisGlnLysGlnValGlu 172
|||||
766 GGAAGCCCAACAGCAGCAGATCCCTCAACCTGAGCAGCTGAAGA 815
172 ySerGlnAsnLysAlaLysGlnCysLysGluSerAlaThrGluAlaGlu 188
|||||
816 AACTCCAAGACAGATAGAAAATGCAACAGGACGTTCTTAAGACCAAG 865
189 ArgValTyrArgGlnAsnIleGluGlnLeuArgAlaArgThrGluTr 205
|||||
866 GACAAGTATGAGAAGTCCCTGAAGGAGCTTGATCAGACCAACCCAGTA 915
205 pGluGlnGluHisArgThrCysGluAlaPheGlnLeuGlnGluPheA 222
|||||
916 CATGGAGAACATGGAGCAGGTGTTTCGAGCAGTGCACGACGTTTGAAGA 965
222 spArgLeuThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeu 238
|||||
966 AGCCCTGCGCTTCTCCGGGAGGTTCTGTGGAGGTTTCAGAACCACTTG 1015
239 SerMetGlnCysValLysAsp...AspGluLeuTyrGluGluValArgLe 254
|||||
1016 GATCTGTCCAATGTGGCTAGCTATAAAACCATTTACCGGAGCTGGAGCA 1065
254 uThrLeuGluGlyCysAspValGluGlyAspIleAsnGlyPheIleGln 271
|||||
1066 GAGCTCANAGCAGCAGATCGGTAGAGGACCTGAGGTGTTCGGGCTA 1115
271 erLysSerThrGly..... 275
|||||
1116 ACCATGGCCAGCATGGCTATGAACTGGCCACAGTTTGAGGAGTGTCT 1165
276 .....ArgGluProProAlaProValPr 283
|||||
1166 GCAGATCTGAATCGAATCTCAGCCGAGAGAGAAAGAGGCTGT.. 1213
283. oTyrGlnAsnTyrTyrAspArgGluValThrProLeuIleGlySerPro 300
|||||
1214 .....GACGTGTCAACCTAACAGGGATCAACC 1241
300 erIleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHis 316
|||||
1242 AGACAGGTGACCATCTGGACAGACAAAG.....CCT 1273
317 GlySerProLysThrThrProSerAlaProAlaAlaSerThr..... 330
|||||
1274 GGCAGCAACCTTAGTGTCCGAGCAACCCGCCAGTCCACGAGGTACA 1323
331 .....GluThrLeuThrProThrP 337
|||||
1324 GTCAGCTACACCCCTTCGAGGAGGAGGACGACACGGGAGCAGCATCA 1373
337 roGluArgAsnGluLeuValTyrAlaSerIleGlu..... 348
|||||
1374 GTGAGAAGGAGGACATTAAAGCCCAAAATCTCAGCAGCTATGAGAAGCT 1423
349 .....ValG 350
1424 CAGACTTACCCACTGACTGGTCTGTATGATGAGTCTAAACAACCCCTTCTC 1473
350 nAlaThrGlnGlyAsnLeuAsnSerSerAlaGlnAsp..... 362
|||||
1474 CTCACGGATGCCACGGGATTTCGAACCCATTTGATGAGGACACGACCT 1523

```

```

363 .....TyrArgAlaLeuTyrAspTyrThrAlaGlnAsn 373
1524 CAGGAACAGAGAGTGGAGCTTCGGGCCCTCTACTGACTATGAGGGCAGGAA 1573
374 SerAspGluLeuAspIleSerAlaGlyAspIleLeuAlaValIleLeuG1 390
1574 CATGATGAGCTGAGCTTCAGGCTGGGATGAACATGACCAAGATAGAGGA 1623
390 uGlyLeuAspGlyTrrpTrpThrValGluArg.....AsnGlyGlnArgG 405
1624 TCAAGATGAACAGGCTTGGTCGAAGGACGTTTAGACACAGCGCCAGGTTG 1673
405 lyPheValproGlySerTyrLeuGluLysLeu 415
1674 GCCTATACCCAGCAACATATGTCGAGGCTATC 1705

seq_name: gb_pri:AB006628

seq_documentation_block:
LOCUS AB006628 2979 bp mRNA PRI 05-FEB-1999
DEFINITION Homo sapiens mRNA for KIAA0290 gene, partial cds.
ACCESSION AB006628
VERSION AB006628.1 GI:2564327
KEYWORDS KIAA0290.
SOURCE Homo sapiens Brain cDNA to mRNA, clone_lib:psport 1 clone:HA6644.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2979)
AUTHORS Ohara,O., Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.
and Nomura,N.
TITLE Direct Submission
JOURNAL Submitted (20-AUG-1997) to the DDBJ/EMBL/GenBank databases. Nobuo
Nomura, Kazusa DNA Research Institute, Gene Structure; 1532-3
Yana, Kizazazu, Chiba 292, Japan (E-mail:cdnainfo@kazusa.or.jp,
URL:http://www.kazusa.or.jp, Tel:+81-438-52-3930,
Fax:+81-438-52-3931)
REFERENCE 2 (bases 1 to 2979)
AUTHORS Ohara,O., Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.
and Nomura,N.
TITLE Prediction of the coding sequences of unidentified human genes
JOURNAL Published Only in DataBase (1997) In press
FEATURES
Location/Qualifiers
1..2979
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HA6644"
/clone_lib="psport 1"
/tissue_type="Brain"
1..2979
/.gene="KIAA0290"
<1..2723
/gene="KIAA0290"
/codon_start=3
/protein_id="BAA22959.1"
/db_xref="GI:2564328"
/translation="LDHWTGPAGVSTETIRMSYFGEHFWGKNGHFEVLYHSVKOGP
ISTKELADIRERATLEETYSKAMKLSKLANGTGMGTFAPLWEFVRYSSDKLALCH
LELTRKLDIKDLRYGEQKTKKKEEVYSTLDADVLSGSQLPKQSRNYL
RCMDORLRESTSQEMDKAETTKKAAEELRRSVVEKYNARSADPEOKMLDGLMPDA
AMEETHLHKMLKASVHEDTHVQIGVHEEFKONTENYVEMLLRKFAESKGTG
REKGGPDLFEAYSAALOAMKRLCAKAFRLPGLSRREPEPPAADFLEPDSTGC
PEVDEGFTVRPDVTONTAEPSRFSRSSDDDEDEPRAFYVHIKPARLQACEPEA
AAQLKATAGSLIIPGGGTGWMKRHSRSDAAGKQPRSPARTSGSCAERLQBEQVSK
NLFPGPLESAFHEDHTGSSILGTSFSPSSPENVEDSGLDSPGSHARPSPDAS
WYPRGTQPSPCRRAPPEARIPPLDPQPLASSPGWLEALAGDLMPPA
DPTARGLAAPPRLRSRVKSCPLTRSGNDLSRSLSPGLSSMASTALERPSFLSQ
GHVCSRGSPVVLGSODALPIATAFTEYHYAVFRGHSPSLCARVYGGELTWTTPAGIVR
VQSGTPPPVLFVLVHTTATIEHFQPNADLLFSDSQSDPETKDFMNNALTEALQAR
QAEQNPTASYINVLLRYTGFSRGPQSPVQLSAHQCCATLTQVSVEGYRPGATAV
PPLTNVQILPLVGEPTVNRQLPATWNLEKRLTWRPLPDVSEAGSGRLSASWEP
SPSTPSPVAAQFTSGTTLSGVDLELVGSGYRMSLVKRRFATGMVLYSC"

```



```

267 ACTGTTTGCACAGGGGCGCATGATCGAAGAGGCGATATGCACAGCAGCTC 316
58 ValGlnIleAlaArgLys.....AlaGlyGlyGlnTh 68
   : : : : : ||||| : :
317 ACTGAGTGGCCCGCCAGCGTGGACAGCTTGTAGAGAAGGACACACAG... 364
   : : : : : ||||| : :
68 rGluMetAsnSerLeuArgThrSerPheAspSerLeuLysGlnInThrG 85
   : : : : : ||||| : :
365 ...TATGGACCGCTGGAGAAGCGCTGATGGCTGCTCATGCTGAGGCAG 410
85 luAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGluGlu 101
   : : : : : ||||| : :
411 AGAGAGTGGGCGCTACACTGGAAAGTGAAGGCATCACTGAACGAA 460
102 LeuArgSerLeuGluPheArgGluArgGlnLysGlu..... 114
   : : : : : ||||| : :
461 .....GACTTTGAGAAGATCAAGAACTGGCAGAAGGAGCCTTTCACAA 504
115 .....GlnArgLysLysTyrGluAlaIleMeta 124
   : : : : : ||||| : :
505 GCAGATGATGGAGGCTTCAAGGAGACCAAGGAAGCAGAGGATGGCTTTC 554
124 spArgValGlnLysSerLysLeuSerLeuTyrLysLysThrMetGluSer 140
   : : : : : ||||| : :
555 GGAAGGCCAGAAAGCCCTGGGCCAAGAGCTGAAGAGGTGGATGCGGCA 604
141 LysLysAlaTyrAspGlnLysCysArgAspAlaAspAla.....G 155
   : : : : : ||||| : :
605 AAGAAACACACACACACAGCTGCAAGAGGAGAGAGCTGGCGCTCTCCG 654
155 uGlnAlaPheGluArgValSerAlaAsnGlyHisGlnLysGlnValGlu 172
   : : : : : ||||| : :
655 GGAAGCCCAACAGCAAGACAGATCCCTCAACCCCTGAGCAGCTGAAGA 704
172 ysSerGlnAsnLysAlaLysGlnCysLysGluSerAlaThrGluAlaGlu 188
   : : : : : ||||| : :
705 AATTGCAGACAGATAGAAAAGTGAAGCAGGAGCTTCTTAAGACCAAG 754
189 ArgValTyrArgGlnAsnIleGluGlnLeuGluArgAlaArgThrGluTr 205
   : : : : : ||||| : :
755 GACAAGTATGAGAAGGCCCTGAGGAGCTTGACCAGACTACACCGCAGTA 804
205 pGluGlnGluHisArgThrThrCysGluAlaPheGlnLeuGlnPheA 222
   : : : : : ||||| : :
805 CATGGAGAACATGCGAGCAGGTGTCGACGAGTGCACGAGTTTGAAGAGA 854
222 spArgLeuThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeu 238
   : : : : : ||||| : :
855 AGCGCCTGGCGCTCTCCGAGAGGTTCTGTGGAGGTTTCAGAGCACTTG 904
239 SerMetGlnCysValLysAspGlu...LeuTyrGluGluValArgLe 254
   : : : : : ||||| : :
905 GACCTGTCCAATGTGGCTAGCTACAAGGCATTTACCGGGAGCTGGAGCA 954
254 uThrLeuGluGlyCysAspValGluGlyAspIleAsnGlyPheIleGln 271
   : : : : : ||||| : :
955 GAGCATCAAGACAGCTGATGCTGTGGAGGACCTGAGGTGTTCCGAGCTA 1004
271 erLysSerThrGlyArgGluProProAlaProValProTyrGlnAsnTyr 287
   : : : : : ||||| : :
1005 ACCATGGCGCAGGCGATGGCTATGNACTGGGCCACAGTTTGGAGGTGTCT 1054
288 TyrAspArgGluValThr..... 293
   : : : : : ||||| : :
1055 GCAGATCTGAATCGAATCTCAGCCGGAAGAGAGAAGAGGCTGCTGA 1104
294 .....ProLeuIleGlySerProSerIleGlnProSerCysGlyValI 308
   : : : : : ||||| : :
1105 TGGTGTCCCTCAACAGGAATCAACCAGACAGAGGTGACCACTGTGGACAGA 1154
308 leLysArgPheSerGlyLeuLeuHisGlySerProLysThrThrProSer 324
   : : : : : ||||| : :
1155 ACAAGCCTAGCAGCACCTT.....AGTGTCCCGAGC 1186

```

```

325 AlaProAlaAlaSerThr..... 330
   : : : : : ||||| : :
1187 AACCCCGCCAGTCCAGCGACTTACAGTCCAGCTACAACCCCTTCGAGGA 1236
331 .....GluThrLeuThrProThrProGluArgAsnGluLeuValTyrA 345
   : : : : : ||||| : :
1237 CGAGGAGCAGACGGGCGAGCGTCACTGTGAGAGGAGGACATTAAAGGCCA 1286
345 laSerIleGlu..... 348
   : : : : : ||||| : :
1287 AAAATGTCAGCAGCTATGAGAAGACTCAGAATTACCCCGCTGACTGTCT 1336
349 .....ValGlnAlaThrGlnGlyAsnLeuAsnse 358
   : : : : : ||||| : :
1337 GATGATGAGTCTAACACCCCTTTCTCTCCACGGATGCCAATGGGACTC 1386
358 rSerAlaGlnAsp.....TyrArgA 365
   : : : : : ||||| : :
1387 TAACCCGTTTGACGAGGACACTACCTCAGGAACAGAAAGTGGAGTTCGAG 1436
365 laLeuTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSerAla 381
   : : : : : ||||| : :
1437 CCCTCTATGACTATGAGGGACAGACACCATGAGCTGAGCTTCAAGGCT 1486
382 GlyAspIleLeuAlaValIleLeuGluGlyGluAspGlyTrpTrpThrVa 398
   : : : : : ||||| : :
1487 GGGATGAGCTGACCAACATAGAGGATGAAGATGAACAGGTTTGGTCAA 1536
398 lGluArg.....AsnGlyGlnArgGlyPheValProGlySerTyrLeuG 413
   : : : : : ||||| : :
1537 GGGACGTTTAGACAGAGTGGCCAGGTGGCTTATACCCAGCAACTATGTG 1586
413 luLysLeu 415
   : : : : : ||||| : :
1587 AGGCTATC 1594

```


PT transportability activity - by fusing candidate DNA with disabled
 PT promoter for transcription factor with reporter gene
 PS Example 1; Pages 34-37; 62pp; Japanese.
 CC The present sequence is used in the course of the invention. The
 CC specification describes a method for the identification of a DNA encoding
 CC a nuclear transport peptide. The method comprises fusing the DNA to DNA
 CC encoding a transcription factor which lacks the ability to transport
 CC nucleic acid, and inserting the fused DNA into a eukaryotic host (such
 CC as a yeast) having in its genome a promoter region activated when the
 CC transcription factor binds to it, and downstream of this promoter a
 CC reporter gene (such as LEU2 or beta-galactosidase). The transformed
 CC eukaryote is then cultured and if the reporter gene is expressed then
 CC the DNA tested encodes a peptide with nuclear transport activity. The
 CC disabled transcription factor is preferably a fusion protein containing
 CC an extranuclear signal sequence, a DNA binding domain (such as LexA) and
 CC a transcription activation domain (such as GAL4). The products can be
 CC used in the investigation and improvement of nucleic acid transcription,
 CC especially in connection with the development of new drugs.
 SQ Sequence 1080 BP; 298 A; 249 C; 268 G; 265 T;

Query Match 16.3%; Score 343; DB 1; Length 1080;
 Best Local Similarity 100.0%; Pred. No. 1.3e-82;
 Matches 343; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 cgcaattttaacaaagtggaattgctgtagctcattgctcctcacttcaactaa 135
 |||||||
 Db 642 CGCAATTTAATCAAGTGGGAATATTGCTGATAGCTCATTTGCTCTTCACTTAA 701
 |||||||

QY 136 cagtagcaacgggtccgaacctatacaactcaaaacttctcaagcgctttcacaacc 195
 |||||||
 Db 702 CAGTAGCAACGGTCCGAACCTCATAACAACCTCAACAAATTCCTCAAGCGCTTTCACAACC 761
 |||||||

QY 196 aattgctcctctaaagttcatgaatactcatgaataatgaataacacggctagtaaat 255
 |||||||
 Db 762 AATTGCTCCTCTAAGCTTCATGATAACTTCATGAATAATGAATAACACGGCTAGTAAAT 821
 |||||||

QY 256 tgatggtgtaataatcaaaacacgtgtcacctggttgacggacacaaactgcgtataa 315
 |||||||
 Db 822 TGATGATGTAATAATCAAAACCACTGTCACCTGGTTGGACGGACCAAACTGCGTATAA 881
 |||||||

QY 316 cgcgtttggaatacactacacgggagtggttataaccactacaatggatgtagtataaacta 375
 |||||||
 Db 882 CGCGTTTGGAAATCACACACAGGATGTTTAATACCACTACAAATGGATGATATATACTA 941
 |||||||

QY 376 tctattcgatgatgaagataccacacacacacacacacacacacacacacacacacacac 418
 |||||||
 Db 942 TCTATTGATGATGAAGATACCCCAACCAACCAACCAACCAACCAACCAACCAACCAAC 984
 |||||||

RESULT 5
 V70966
 ID V70966 standard; cDNA to mRNA; 342 BP.
 AC V70966;
 DT 04-MAR-1999 (first entry)
 DE Transcription activation domain GAL4.
 KW Nuclear transport peptide; transcription factor; DNA binding domain;
 KW LexA; nuclear transport activity; transcription activation domain; GAL4;
 KW nucleic acid transcription; ss.
 OS Unidentified.
 PN WO9849284-A1.
 PD 05-NOV-1998.
 PF 27-APR-1998; J01936.
 PR 24-OCT-1997; JP-309686.
 PR 28-APR-1997; JP-124795.
 PA (HELI-) HELIX RES INST.
 PI Noguchi T, Ueki N;
 DR WPI; 99-024053/02.
 DR P-PSDB; W85000.
 PT Detection and isolation of DNA encoding peptides with nuclear
 PT transportability activity - by fusing candidate DNA with disabled
 PT transcription factor DNA and introducing into eukaryote with

PT promoter for transcription factor with reporter gene
 PS Example 1; Pages 30-31; 62pp; Japanese.
 CC The present sequence encodes transcription activation domain
 CC GAL4, which is used in the course of the invention. The specification
 CC describes a method for the identification of a DNA encoding
 CC a nuclear transport peptide. The method comprises fusing the DNA to DNA
 CC encoding a transcription factor which lacks the ability to transport
 CC nucleic acid, and inserting the fused DNA into a eukaryotic host (such
 CC as a yeast) having in its genome a promoter region activated when the
 CC transcription factor binds to it, and downstream of this promoter a
 CC reporter gene (such as LEU2 or beta-galactosidase). The transformed
 CC eukaryote is then cultured and if the reporter gene is expressed then
 CC the DNA tested encodes a peptide with nuclear transport activity. The
 CC disabled transcription factor is preferably a fusion protein containing
 CC an extranuclear signal sequence, a DNA binding domain (such as LexA) and
 CC a transcription activation domain (such as GAL4). The products can be
 CC used in the investigation and improvement of nucleic acid transcription,
 CC especially in connection with the development of new drugs.
 SQ Sequence 342 BP; 120 A; 77 C; 55 G; 90 T;

Query Match 16.3%; Score 342; DB 1; Length 342;
 Best Local Similarity 100.0%; Pred. No. 1.5e-82;
 Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 gcaaatTTTaaTcaaaagtggaattgctgtagctcattgctcctcacttcaactaa 136
 |||||||
 Db 1 GCCAATTTAATCAAGTGGGAATATTGCTGATAGCTCATTTGCTCTTCACTTAAAC 60
 |||||||

QY 137 agtagcaacgggtccgaacctatacaactcaaaacttctcaagcgctttcacaacca 196
 |||||||
 Db 61 AGTAGCAACGGTCCGAACCTCATAACAACCTCAACAAATTCCTCAAGCGCTTTCACAACCA 120
 |||||||

QY 197 attgctcctctaaagttcatgaataactcatgaataatgaataacacggctagtaaat 256
 |||||||
 Db 121 AATGCTCCTCTAAGCTTCATGATAACTTCATGAATAATGAATAACACGGCTAGTAAAT 180
 |||||||

QY 257 gatggtgtaataatcaaaacacgtgtcacctggttgacggacacaaactgcgtataac 316
 |||||||
 Db 181 GATGATGTAATAATCAAAACCACTGTCACCTGGTTGGACGGACCAAACTGCGTATAAC 240
 |||||||

QY 317 ggcgtttggaatacactacacgggagtggttataaccactacaatggatgtagtataaacta 376
 |||||||
 Db 241 CGCGTTTGGAAATCACACACAGGATGTTTAATACCACTACAAATGGATGATATATACTAT 300
 |||||||

QY 377 ctattcgatgatgaagataccacacacacacacacacacacacacacacacacacacac 418
 |||||||
 Db 301 CTATTGATGATGAAGATACCCCAACCAACCAACCAACCAACCAACCAACCAACCAAC 342
 |||||||

RESULT 6
 Q61607
 ID Q61607 standard; DNA; 2646 BP.
 AC Q61607;
 DT 24-NOV-1994 (first entry)
 DE Mutated GAL4 gene coding for Asp insertion between Asp863 and Val864.
 KW Mutated GAL4; heterologous gene expression; enhanced; increased;
 KW protein production; positive regulator; transcription activator;
 KW galactose metabolism; ss.
 OS Eukaryota.
 PH Key
 FT mat_peptide 1..2646
 FT /*tag= a
 FT /product= mutated_GAL4
 FT /transl_except= (pos:190..192, aa:Lys)
 FT /transl_except= (pos:757..759, aa:Ser)
 FT /transl_except= (pos:1375..1377, aa:Glu)
 FT /transl_except= (pos:1990..1992, aa:Cys)
 PN J06078767-A.
 PD 22-MAR-1994.
 PF 07-SEP-1992; 262723.
 PR 07-SEP-1992; JP-262723.
 PA (GREC) GREEN CROSS CORP.

[illegible]

```

RESULT      7
ID          Q61611
AC          Q61611 standard; DNA; 107 BP.
DT          24-NOV-1994 (first entry)
DE          Wild-type GAL4 C-terminal coding sequence.
KW          Mutated GAL4; heterologous gene expression; enhanced; increased;
KW          protein production; positive regulator; transcription activator;
KW          galactose metabolism; ds.
OS          Synthetic.
key         Location/Qualifiers
FH          key
FT          misc_feature      2..103
FT          FT               /*tag= a
FT          FT               /*note= "encodes amino acids 848-881 of wild-type
FT          FT               GAL4"
FT          FT               1..4
FT          FT               /*tag= b
FT          FT               /*label= sticky_end
FT          FT               /*note= "the complementary strand overhangs Q61611
FT          FT               by the 4 base sequence 5'-TCCG-3'."

```

PN J06078767-A.
PD 22-MAR-1994.
PE 07-SEP-1992; 262723.
PF 07-SEP-1992; JP-262723.
PG (GREC) GREEN CROSS CORP.
PH WPI: 94-131274/16.
PI P-PSDB: R62037.
PJ Mutated GAL4 and a method for the expression of foreign protein -
PK useful for enhancing expression of protein
PL Example 1; Page 14; 20pp; Japanese.
PM A mutant GAL4 in which an Asp residue is inserted between wild-type
PN Asp863 and Val1864 is claimed. The mutated GAL4 can be used for

CC	increasing the expression of foreign genes under the control of
CC	the GAL1, GAL7 or GAL10 promoters. Sequences coding for the
CC	C-terminal amino acids of wild-type and mutant GAL4 (Q61611 and
CC	Q61613, respectively) were each synthesised from two overlapping
CC	complementary oligonucleotides.
SQ	Sequence 107 BP; 41 A; 21 C; 19 G; 26 T;
Query Match 4.9%; Score 103; DB 1; Length 107;	
Best Local Similarity 100.0%; Pred. No. 1.2e-18;	
Matches 103; Conservative 0; Mismatches 0; Indels 0;	
Qy	316 cgcgatttggaatcactacagggaattttaatcaccactacaatggatgatgtataaacct
Db	
Db	1 CGCGTTTGGAAATCAGTACAGGGATGTTTAAATACCACTACAAATGGATGATATATAACT
Qy	376 tctattcgatgatgaagatacccccacaaaccccaaaaaaagag 418
Db	
Db	61 TCTATTGCATGATGAGATACCCCAACCAACCCAAAAAAGAG 103
RESULT	8
ID	Q61613 standard; DNA; 110 BP.
AC	Q61613; 1994 (first entry)
DE	Mutated GAL4 C-terminal coding sequence.
KW	Mutated GAL4; heterologous gene expression; enhanced; increased;
KW	protein production; positive regulator; transcription activator;
KW	galactose metabolism; ds.
OS	Synthetic.
FH	Key
FT	misc_feature 2..106 Location/Qualifiers
FT	/tag= a
FT	/note= "Encodes amino acids 848-881 of wild-type
FT	GAL4 with additional Asp residue inserted
FT	between Asp 863 and Val 864"
FT	1..4
FT	misc_feature /tag= b
FT	/label= sticky_end
FT	/note= "the complementary strand overhangs Q61611
FT	by the 4 base sequence 5'-TCGA-3'"
PN	J06078767-A.
PD	22-MAR-1994.
PF	07-SEP-1992; 262723.
PR	07-SEP-1992; JP-262723.
PA	(GREC) GREEN CROSS CORP.
DR	WPI; 94-131274/16.
DR	P-PSDB: R62038.
PT	Mutated GAL4 and a method for the expression of foreign protein
PT	useful for enhancing expression of protein
PS	Example 1; Page 15; 20pp; Japanese.
CC	A mutant GAL4 in which an Asp residue is inserted between wild-t-
CC	Asp863 and Val864 is claimed. The mutated GAL4 can be used for
CC	increasing the expression of foreign genes under the control of
CC	the GAL1, GAL7 or GAL10 promoters. Sequences coding for the
CC	C-terminal amino acids of wild-type and mutant GAL4 (Q61611 and
CC	Q61613, respectively) were each synthesised from two overlapping
CC	complementary oligonucleotides.
SQ	Sequence 110 BP; 42 A; 21 C; 20 G; 27T;

	Query Match	4.3%	Score 90;	DB 1;	Length 110;
	Best Local Similarity	97.2%;	Pred. No. 3.9e-15;		
	Matches 103;	Conservative 0;	Mismatches 0;	Indels 3;	Gaps 1;
Qy	316	cgcgttggaaatcaactacagcaggagtgtttaataccaccatacaatg---gatgatgtatataa	372		
Dd	1	CGCGTTTGGAACTACTACAGGATGTTTTAATACCACCTACAATGGATGATGATGTATATAA	60		
Qy	373	ctatctattcgatgatgaagataccccaccacccccaaaaaaaagag	418		
Dd	61	CTATCTATTTCGATGATGAAGATACCCACCAAAAGCCCCAAAAAACAG	106		

```

RESULT 9
ID V21683
AC V21683; standard; DNA; 9600 BP.
DT 17-AUG-1998 (first entry)
DE Vector plasmid pCMVkmrR-EPI.
KW Polynucleotide delivery; plasmid pCMVkmrR-EPI; vector;
KW gene therapy; vaccine; polycationic agent; ss.
OS Chimeric - Epstein-Barr virus.
OS Chimeric - Adeno-associated virus.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Bos taurus.
FH Key
FT Location/Qualifiers
FT 14..2594
FT /tag= a
FT /product= "EBV nuclear antigen A"
FT misc_feature
FT 2623..4559
FT /tag= b
FT /note= "EBV origin of replication"
FT repeat_unit
FT 4928..5104
FT /tag= c
FT /rpt_type= INVERTED
FT /note= "AAV inverted terminal repeat"
FT repeat_unit
FT 7189..7355
FT /tag= d
FT /rpt_type= INVERTED
FT /note= "AAV inverted terminal repeat"
FT promoter
FT 5112..6734
FT /tag= e
FT /note= "CMV immediate-early enhancer/promoter"
FT terminator
FT 6818..7050
FT /tag= f
FT /note= "bovine growth hormone polyA sequence"
PN W09806437-A2.
PD 19-FEB-1998.
PE 13-AUG-1997; U14465.
PF 13-AUG-1996; US-023867.
PI (CHIR ) CHIRON CORP.
PI Cohen F, Dubois-Stringfellow N, Dwarki V, Innis MA,
PI Murphy JE, Tetsuo U, Zukermann R;
PI WPI; 98-159296/14.
DR Polycationic agents based on alpha-amino acids, able to complex
PT with nucleic acid - to facilitate its entry into cell, condense it
PT and protect it against serum degradation, particularly for use in
PT gene therapy
PT Disclosure; Page 77-80; 100pp; English.
PS This polynucleotide comprises the DNA sequence of vector plasmid
CC pCMVkmrR-EPI, which contains an Epstein-Barr virus (EBV) origin
CC of replication from plasmid pCEP4, a coding region for EBV nuclear
CC antigen A from pCEP4, a pair of inverted terminal repeats from
CC adeno-associated virus, a cytomegalovirus enhancer/promoter, a
CC bovine growth hormone polyA sequence, and a kanamycin resistance
CC selectable marker. Polynucleotides encoding polypeptides, such as
CC erythropoietin or leptin, and ribozymes and antisense
CC polynucleotides can be inserted into the vector. The vector is
CC preferred for use in novel compositions and methods for improved
CC polynucleotide delivery into cells. In these methods, polycationic
CC agents are used to increase the frequency of uptake of a
CC nucleic acid (see also V21684-86) into a cell. The polycationic
CC agent can condense with the nucleic acid and inhibit serum and/or
CC nuclease degradation of the nucleic acid. The nucleic acid can be
CC a vector, may express a therapeutic protein or a vaccinating viral
CC or cancer antigen, or is itself therapeutic (antisense or
CC ribozyme). The methods and compositions can be used in the gene
CC therapy of many diseases.
SQ Sequence 9600 BP; 2326 A; 2376 C; 2817 G; 2081 T;

```

Query Match 3.0%; Score 63.4; DB 1; Length 9600;
 Best Local Similarity 43.7%; Pred No. 4e-07;
 Matches 329; Conservative 0; Mismatches 421; Indels 3; Gaps 1;

```

QY 701 tccgagatgccttctgtgcagggaacttcacggccacacaggggtatgaggtgtactgc 760
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 657 TCCAAGTTGCATTTGGCTGCAAAAGGGACCACGGTGTGAACAGGAGCAGGAGCGGG 716
QY 761 agaggtctgtgcacggcaggaagatgtcaaggatgtggaggagctgctcagacagagg 820
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 717 AGGGCAGGAGCAGGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 776
QY 821 ccaggcggagagaggtacgggaagagctggtgagattgcacacaaaggtcgtggtgccc 880
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 777 AGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 836
QY 881 agacagagatgaattcccttgaggacacctcttgaactccctgaagcagcaaacagagaatg 940
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 837 AGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 896
QY 941 tgggcagtgacacatccagctgcccctgctgctgctgctgctgctgctgctgctgctgctg 1000
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 897 GGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 956
QY 1001 agttccgagagacagaaagagcagcgaagaagtatgagggccatcatgagacctgtcc 1060
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 957 AGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1016
QY 1061 agaagagcaagtgtgcctctcaagaagaccatggagtcctcaagaagggcatatgaccaga 1120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1017 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1076
QY 1121 agtcaggatgcagatgatctgtagcaggccttcagtcgctgtagtcgaatggccacc 1180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1077 AGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1136
QY 1181 agaagcaagtgaagagagccagaaacaaagcagtcgaaggagtcagccacagagg 1240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1137 AGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1193
QY 1241 cagaagagtgtagcagcaaatatcgacaactggagagagcgagccgagtcggtgagc 1300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1194 AGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1253
QY 1301 aggagcaccggactacctgtgaggccttcagtcaggagtttgaccggtccaccatcc 1360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1254 AGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1313
QY 1361 tccgcaatgccctgtggtgcaactgtaaccagctctccatgcaagtgtcgaagatgatg 1420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1314 AGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1373
QY 1421 agctctatgaggaagtgcggtgaccttgagg 1453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1374 AGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1406

RESULT 10
Q51731
ID Q51731 standard; DNA; 10596 BP.
AC Q51731;
DT 31-MAY-1994 (first entry)
DE Plasmid pCISBSON for subcloning huHGF variants.
KW Hepatocyte Growth Factor; HGF; variant; mutein; in vitro mutagenesis;
KW proteolysis resistant; liver; malignancy; CMV-driven;
KW Cytomegalovirus; episomal expression plasmid; ss.
OS Synthetic.
FH Key
FT Location/Qualifiers
FT enhancer
FT 1..611
FT /tag= a
FT /note= "CMV enhancer/promoter"
FT promoter
FT 758..775
FT /tag= b
FT /label= SP6_promoter
FT misc_feature
FT 845..849
FT /tag= c

```



```

FT misc_feature /note= "SP6 RNA start"
FT 902..966
FT /*tag= d
FT /function= cloning_linker
FT poly_a_signal 967..1107
FT /*tag= e
FT /note= "SV40 poly A"
FT 1108..1531
FT /*tag= f
FT /function= SV40_origin
FT misc_feature 1580..4189
FT /*tag= g
FT /label= EBNA-1
FT 4190..6374
FT /*tag= h
FT /function= orip
FT misc_feature 4295..4887
FT /*tag= i
FT /note= "family of repeats"
FT 5866..5978
FT /*tag= j
FT /note= "dyad region"
FT 6375..6457
FT /*tag= k
FT /label= HSV_TK_terminator_3'-end
FT 6975..7975
FT /*tag= l
FT /phenotype= neomycin_resistance
FT /note= "Tp5 neomycin phosphotransferase gene"
FT 7975..8112
FT /*tag= m
FT /label= TK_promoter
FT 8114..8594
FT /*tag= n
FT /function= M13_ori
FT 8595..10414
FT /*tag= o
FT /label= delta_2a
FT
FT W09323541-A.
FT 25-NOV-1993.
FT 17-MAY-1993; U04648.
FT 18-MAY-1992; US-884811.
FT 18-MAY-1992; US-885971.
FT (GETH ) GENENTECH INC.
FT Godowski PJ, Lokker NA, Mark MR;
FT WPI; 93-386573/48.
FT Hepatocyte growth factor variants - are resistant to proteolytic
FT cleavage into its two-chain form, used to treat malignancies
FT associated with HGF receptor
FT Example 1; Fig 6; 87pp; English.
FT Plasmid pcISEBON (a PRK5 derivative) is an episomal CMV driven
FT expression plasmid. HUHGF variants with enhanced receptor binding
FT activity were produced by site-directed mutagenesis. Stable
FT populations of preferred HGF variants were obtained by transfecting
FT human embryonic kidney 293 cells and then these were subcloned in
FT pcISEBON. See R52940-R52949 for examples of pref. HGF variants.
FT Sequence 10596 BP; 2625 A; 2571 C; 3024 G; 2376 T;
FT
FT Query Match 3.08; Score 63.4; DB 1; Length 10596;
FT Best Local Similarity 43.7%; Pred. No. 4.2e-07;
FT Matches 329; Conservative 0; Mismatches 421; Indels 3; Gaps 1;
FT
FT 701 tccgagatgccttctgtgagcagacttcacgcccacacaggtatgagtgctactgc 760
FT ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 2191 TCCAGTGTGCTATGGCTGCAAGGACCCACGCTGGACAGAGCAGGAGCGCGG 2250
FT
FT 761 agagcgtctgagcagcaggaagtgtgcaagagtgaggagctgctcagacagaggg 820
FT ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 2251 AGGGCAGGACGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2310
FT ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 821 cccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 880
FT ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 2311 AGGGCAGGAGGGGCGAGGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2370
Qy 881 agacagagatgaattccctgagggacctcttctgactccctgaagcagcaaacagagaa 940
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2371 AGGAGGGGCGAGGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2430
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 941 tgggcagtgacacatccagctgcccctgcccctgctgctgctgctgctgctgctgctg 1000
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2431 GGCAGGAGCAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGG 2490
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1001 agttccgagagacacagaaagcagcggaagaaagaaagaaagaaagaaagaaagaa 1060
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2491 AGCAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGG 2550
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1061 agaagagcaagtgtcgtctctacaagaagaccatgagttccaagaagcagcagcagcaga 1120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2551 AGGAGCAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGG 2610
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1121 agtcagggatgcagatgatgctgagcagggccttcgagcgtgtgagtgccaatgcccacc 1180
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2611 AGGGCAGGAGCAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGG 2670
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1181 agaagcaagttagaaaagagccagaaacaaagccaaagcagtcgaagcagtcagccacagagg 1240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2671 AGCAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGG 2727
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1241 cagaaagagtgctacagggcaaaataatcgcaacactgagagagcagcagcagcagcagcagc 1300
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2728 AGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGG 2787
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1301 aggagcaccgactacctgtgagccttcaggttccagagtttgacgggttcaccatcc 1360
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2788 AGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGG 2847
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1361 tccgcaatgccctgtggtgcaactgtaaccagctctccatcagctgtcgaagatgatg 1420
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2848 AGGAGCAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGG 2907
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1421 agctctatgaggaagtcgctgacctgaccttgagg 1453
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2908 AGGGCAGGAGCAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGG 2940
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 11

```

T40348
ID T40348 standard; DNA; 10596 BP.
AC T40348;
DT 09-DEC-1996 (first entry)
DE Plasmid pcISEBON for expression of hepatocyte growth factor.
KW Human; hepatocyte growth factor; HGF; huHGF; serum; proteolytic cleavage;
KW pro-hormone; beta subunit; alpha subunit; kringle domain; prothrombin;
KW plasminogen; catalytic domain; serine protease; HGF variant;
KW HGF receptor; malignancy; chronic HGF receptor activation; ss.
OS Synthetic.
PN US5547856-A.
PD 20-AUG-1996.
PF 18-MAY-1992; 884811.
PF 18-MAY-1992; US-885971.
PR 18-MAY-1992; US-884811.
PR 13-JUL-1993; US-087783.
PA (GETH ) GENENTECH INC.
PI Godowski PJ, Lokker NA, Mark MR;
DR WPI; 96-392634/39.
PT New hepatocyte growth factor variants - are resistant to in vivo
PT proteolytic cleavage into a 2-chain form, useful as HGF antagonists
PS Example 1; Fig 6; 39pp; English.
CC This sequence represents the episomal CMV driven expression plasmid
CC pcISEBON which was used in the expression of variant human hepatocyte
CC growth factor (HGF). HGF is isolated from human serum and is a
CC disulphide linked heterodimer derived by proteolytic cleavage of the
CC pro-hormone between residues 494 and 495. This generates a molecule
CC composed of an alpha subunit of 440 amino acids (mol. wt. 69 kD) and
CC a beta subunit of 234 amino acids (mol. wt. 34 kD). The alpha and beta

```


CC formula [(Glya)X(Glyb)Y(Glyc)Z]n where Glya, Glyb, Glyc are 1-6
CC sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,
CC Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not
CC be identical from n repeat to n repeat. Alternatively a nucleic acid
CC encoding the stabilising polypeptide can be linked onto or inserted into
CC a nucleic acid encoding a core protein. The fusion proteins of the
CC invention are more resistant to degradation by proteases and, thus, have
CC a longer half-life than the unfused core protein. The products can be
CC used for treating autoimmune diseases, cancer and inflammation. In
CC particular, the core protein may be an IkappaB regulator protein for the
CC treatment of inflammatory bowel disease, or a nitroreductase protein
CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer,
CC or other pathological conditions. The fusion proteins can also be used in
CC diagnostic methods such as in vivo imaging.

SQ Sequence 799 BP; 201 A; 106 C; 479 G; 13 T;

Search completed: September 15, 2000, 10:07:01
Job time: 2280 sec

Query Match		2.88;	Score 59.4;	DB 1;	Length 799;
Best Local Similarity		43.68;	Pred. No. 1.6e-06;		
Matches	316;	Conservative	0;	Mismatches	406;
				Indels	3;
				Gaps	1;
Qy	729	cacggccacacacaggtatgagtgctactgtcagaggtgctgtgacggtcaggaagatgtg	788		
Db	14	CACGGTGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	73		
Qy	789	caaggatgttgaggagctgtctcagacagagggcccgagggagaggtacgsgaaagga	848		
Db	74	GGAGGAGGGGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA	133		
Qy	849	gctgtgtcagattgcacgcaagctgtgtgccagacagagatgaattccctgaggaacctc	908		
Db	134	GCAGGAGGGGGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA	193		
Qy	909	ctttgactccctgaagcagcaaacagagaaatgtgtggcagtgccacacatccagctggccct	968		
Db	194	GGAGGGGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGG	253		
Qy	969	ggccctgcgtgagagagctgcggagccctgagaggttcccgagagagacagaaagagcagcg	1028		
Db	254	GCAGGAGGGGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA	313		
Qy	1029	gaagaagtatgagcccatcatgaccctgtccagaagcaagtgtcgtctctacaagaa	1088		
Db	314	GCAGGAGGGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA	373		
Qy	1089	gaccatggagtccaagaaggcatatgaccagaagtgcaggatgcagatgatgctgagca	1148		
Db	374	GGGCGAGGAGGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA	433		
Qy	1149	ggcccttcagcgtgtgagtgccaatggccaccagaagaagtagaataagccagaacaa	1208		
Db	434	GGGCGAGGAGGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA	493		
Qy	1209	agccaacagtcgaagagtcagccacacagagcagaagaagtgtacaggcaaaataatcga	1268		
Db	494	GGAGCAGGAGGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA	550		
Qy	1269	acaactggagagcggagggaccagtggtggagcagagcaccgggactacctgtgagggcctt	1328		
Db	551	GGGCGAGGAGGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA	610		
Qy	1329	ccagttgcagagagtttgaccgggtccacatctcctccgaatgcctgtggtgcactgttaa	1388		
Db	611	GGAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA	670		
Qy	1389	ccagctctccatcagatgtgtcaaggtatgatgaactctatgaggaagtgcggtgacctt	1448		
Db	671	GCAGGAGGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA	730		
Qy	1449	tgagg	1453		
Db	731	GGAGG	735		

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

L3 ANSWER 2212 OF 2212 EMBASE COPYRIGHT 2000 ELSEVIER SCI. B.V.
AN 84156159 EMBASE
DN 1984156159
TI Evidence for the platelet-derived growth factor-stimulated tyrosine phosphorylation of the platelet-derived growth factor receptor in vivo. Immunopurification using a monoclonal antibody to phosphotyrosine.
AU Frackelton Jr. A.R.; Tremble P.M.; Williams L.T.
CS Howard Hughes Medical Institute, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, United States
SO Journal of Biological Chemistry, (1984) 259/12 (7909-7915).
CODEN: JBCHA3
CY United States
DT Journal
FS 029 Clinical Biochemistry
016 Cancer
025 Hematology
026 Immunology, Serology and Transplantation
LA English
AB The addition of platelet-derived growth factor (PDGF) to intact BALB/c
3T3

cells results in the rapid (<1 min), dose-dependent phosphorylation of a number of proteins that could be isolated by a monoclonal **anti-phosphotyrosine** antibody. The predominant tyrosine-phosphorylated protein shared many characteristics with the PDGF receptor, including its molecular weight (170,000), isoelectric point (pI of about 4.2), its binding to DEAE-cellulose, and its pattern of binding to lectins. This 170-kDa protein, labeled with [35S]methionine, was substantially purified from PDGF-stimulated cells using the monoclonal **anti-phosphotyrosine** antibody but was not significantly immunopurified from unstimulated cells. At 37.degree.C, phosphorylation of the 170-kDa protein was maximal by 5-10 min of exposure to PDGF, and thereafter decreased rapidly. However, at 4.degree. C, the phosphorylation continued to increase after 3 h of exposure to PDGF. Subsequently, shifting the cells from 4 to 37.degree.C resulted in an additional rapid burst of tyrosine phosphorylation. Among the other PDGF-stimulated molecules, the most prominent and consistently observed was a cytosolic, acidic (pI of about 4.2), 74-kDa protein. These findings indicate that the action of PDGF in vivo is associated with the rapid and transient tyrosine phosphorylation of several membrane and cytosolic proteins; the most prominent of these proteins, isolated by monoclonal antibody to phosphotyrosine, is likely to be the PDGF receptor. The use of this antibody provides a new approach for purification of the PDGF receptor.

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using SW model

Run on: September 15, 2000, 09:29:01 ; Search time 80.09 Seconds
(without alignments)

(without alignments)
6560.162 Million cell updates/sec

Title: US-09-068-377-2

Perfect score:	2100	2100
Sequence:	1 caataattcaagctatacca.....gaagyaaaaaaaaaa	

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match 0%
	Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2100	100.0	2100	1	V57973	Murine tyrosine ph
2	346.6	16.5	7625	1	T71323	Hybrid activation
3	343.8	16.4	980	1	T43136	GAL4 insert in in
4	343	16.3	1080	1	V70970	Nucleotide sequenc
5	342	16.3	342	1	V70966	Transcription accl
6	332.4	15.8	2646	1	O61607	Mutated GAL4 gene
7	103	4.9	107	1	O61611	Wild-type GAL4 C-t
8	90	4.3	110	1	O61613	Mutated GAL4 C-ter
9	63.4	3.0	9600	1	V21683	Vector plasmid PCM
10	63.4	3.0	10596	1	O51731	Plasmid pCISBON f
11	63.4	3.0	10596	1	T40348	Plasmid pCISBON f
12	63.4	3.0	10596	1	X15650	Nucleotide sequenc
13	61.6	2.9	32207	1	V73805	KSHV LTR DNA (nucl
14	61.6	2.9	137550	1	V19941	KSHV long unique c
15	59.4	2.8	799	1	V55831	Nucleotide sequenc
16	59	2.8	1039	1	V18439	Unknown DNA sequen
17	57.2	2.7	2004	1	T83356	Human SH3P7 gene.
18	56.4	2.7	1025	1	V18441	Gene encoding plas
19	52	2.5	1389	1	T39795	FLCA insert stabl
20	50.8	2.4	1908	1	N71064	Unknown DNA sequen
21	50.6	2.4	795	1	V55830	Plasmodium vivax c
22	50.6	2.4	1042	1	V18442	Mouse SH3p3 gene.
23	50	2.4	1137	1	Q33061	N. clavipes draglin
24	48.8	2.3	835	1	T39787	N. clavipes draglin
25	48.8	2.3	2338	1	V33249	Gene encoding plas
26	48.8	2.3	2338	1	N71065	AMNL chromosome in
27	48.4	2.3	2000	1	N71065	Abi interactor pro
28	47.8	2.2	2887	1	T66904	Nephila clavipes s
29	47.2	2.2	1930	1	Q14184	Nephila clavipes s
30	46.6	2.2	1995	1	V32250	AMNL chromosome in
31	46.6	2.2	2680	1	O84590	CDNA clone LS02-36
32	45.6	2.2	1045	1	T79835	
33	45.6	2.2	1045	1	T79835	

45	4.3	8	2.1	2003	1	T39807	Mouse Hsl gene. Id
44	4.3	8	2.1	1773	1	O38659	Morphogen hox2 cod
43	4.3	8	2.1	1733	1	O28737	Human osteogenic p
42	4.5	2	2.2	3143	1	V18892	CD2 associated int
41	4.5	2	2.2	1659	1	V63342	Ym06 cDNA clone en
40	4.5	2	2.2	1659	1	T87045	CD2 associated int
39	4.5	2	2.2	1592	1	T79833	CDNA clone Ym06 en
38	4.5	2	2.2	1392	1	LS02-21	CD2 associated int
37	4.5	2	2.2	1392	1	V63343	LS02-21 cDNA clone
36	4.5	2	2.2	1392	1	T79834	CDNA clone LS02-21
35	4.5	6	2.2	1045	1	V63344	CDNA clone LS02-21
34	4.5	6	2.2	1045	1	T87047	CD2 associated int

ALIGNMENTS

RESULT	1
ID	V57973 standard; cDNA; 2100 BP.
AC	V57973.
DE	24-NOV-1998 (first entry)
DE	Murine tyrosine phosphorylated cleavage furrow-associated protein cDNA.
KW	Mouse; tyrosine phosphorylated cleavage furrow-associated protein;
KW	PSRIP; PST family; protein tyrosine phosphatase; murine;
KW	polymerisation; actin monomer; eukaryotic cell; identification;
KW	antagonist; ss.
OS	Mus sp.
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	/tag="a"
FT	/product="PSRIP"
FT	/note="tyrosine phosphorylated cleavage furrow-associated protein"
FT	
PN	MO9835037-A1.
PD	13-AUG-1998.
PD	30-JAN-1998; U01774.
PF	29-SEP-1997; U5-938829.
PR	07-FEB-1997; U5-798419.
PA	(SETH) GENENTECH INC.
PI	Dowbenko DJ, Lasky LA;
DR	WPI: 98-447234/38.
DR	P-PSDB: W71595.
PT	New PST-type protein tyrosine phosphatase interacting polypeptide -
PT	nucleic acids and vectors; for inducing the polymerisation of actin
PT	monomers in eukaryotic cells and identifying antagonists
PS	Example 9; Page 62-63; 111P; English.
CC	The present sequence encodes murine tyrosine phosphorylated cleavage
CC	furrow-associated protein (PSRIP), which is a PST-type protein
CC	tyrosine phosphatase (PTP)-interacting polypeptide. PSRIP induces the
CC	a vector containing the nucleic acid sequence encoding PSRIP into the
CC	cell. Assays for identifying (ant)agonists of PSRIP comprise contacting
CC	PSRIP with the agent and monitoring the ability of PSRIP to induce
CC	actin polymerisation.
SO	Sequence 2100 BP; 548 A; 540 C; 594 G; 418 T;
Query Match	100.0%; Score 2100; DB 1; Length 2100;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2100; Conservative	0; Mismatches 0; Indels 0; Gaps
OY	1 caatattcaagcatatcacaagcatacaatcaactccaagctatgccccagaagaagc 60
DB	1 CAAATATTTCAGACTATACCAAGCATCAATCAACTCCAAAGCTTATGCCAAGAAAGACG 60
OY	61 gaaggtctcgaagggcgccaattcaatcaaaatggaatgcatgcatgcatgttc 120
DB	61 GAAGGCTTCGAGGCGGCCCAATTTTATATCAAAAGTGGAAATATGATGAGCTCATTTGTC 120
OY	121 cttcaacttcaactacagtagcaacggtccgaacctatacaacttaaacacttaatttca 180
DB	121 CTTCACCTTTACCTAAAGTAGCAACGGTCCGAAACCTCATTAACAACATCAACAAATTTCTCA 180

THIS PAGE BLANK (USPTO)

[illegible]

RESULT	A
ID	V70970 standard; CDNA to mRNA; 1080 BP.
AC	V70970;
DT	04-MAR-1999 (first entry)
DE	Nucleotide sequence of the specification.
KW	Nuclear transport peptide; transcription factor; DNA binding domain; GLA
KM	LexA; nuclear transport activity; transcription activation domain; GLA
OS	nucleic acid transcription; ds.
FH	Unidentified.
FT	Key Location/Qualifiers
CDS	1..1080
	/*tag= a
PN	WO9849284-A1.
PD	05-NOV-1998.
PF	27-APR-1998; J01936.
PR	24-OCT-1997; JP-309686.
PA	28-APR-1997; JP-124795.
PB	(HELI-) HELIX RES INST.
PI	Noguchi T, Ueki N;
DR	WPI: 99-024053/02.
PT	Detection and isolation of DNA encoding peptides with nuclear

THIS PAGE BLANK (USPTO)

transportability activity - by fusing candidate DNA with disabled promoter for transcription factor with reporter gene

Example 1: Pages 34-37: 62pp; Japanese.

The present sequence encodes a method for the identification of a DNA encoding a nuclear transport peptide. The method comprises fusing the DNA to DNA encoding a transcription factor which lacks the ability to transport a nuclear acid, and inserting the promoter region activated when the nuclear acid, having in its genome a promoter region of this promoter a transcription factor binds to it, and downstream of this promoter a reporter gene (such as LEU2 or beta-galactosidase). The transformed eukaryote is then cultured and if the reporter gene is expressed then the DNA tested encodes a peptide with nuclear transport activity. The disabled transcription factor is preferably a fusion protein containing an extranuclear signal sequence, a DNA binding domain (such as LexA) and a transcription activation domain (such as GAL4). The products can be used in the investigation and improvement of nucleic acid transcription, especially in connection with the development of new drugs.

Query Match 16.3%; Score 343; DB 1; Length 1080;
Best Local Similarity 100.0%; Pred. No. 1.3e-82; Indels 0; Gaps 0;
Matches 343; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query 76 ggcgaatttaacaaagtggaataattgctgtagctcattgctcctcactactaa 135
DB 642 ccccaatttttaacaaagtggaataattgctgtagctcattgctcctcactactaa 195
QY 136 cagtagcaaggtccgaacctcaatacaactcaacaattctcaagcgtttcaaac 761
DB 702 cactacacacggtccgaacctcaatacaactcaacaattctcaagcgtttcaaac 255
QY 196 aattgctcctcactacgttcatactcaactcaataataataacacggttagtaaat 821
DB 762 aattgctcctcactacgttcatactcaactcaataataataacacggttagtaaat 315
QY 256 tgatgatgataataatcaaaacacctgacacctgttgagcagcaaacctggtat 881
DB 822 tgatgatgataataatcaaaacacctgacacctgttgagcagcaaacctggtat 375
QY 316 cgcgttgatgataataatcaaaacacctgacacctgttgagcagcaaacctggtat 941
DB 882 cgcgttgatgataataatcaaaacacctgacacctgttgagcagcaaacctggtat 418
QY 376 tctatcgtatgataataatcaaaacacctgacacctgttgagcagcaaacctggtat 984
DB 942 tctatcgtatgataataatcaaaacacctgacacctgttgagcagcaaacctggtat

RESULT 5
ID V70966 standard: cDNA to mRNA; 342 BP.
AC V70966:
DT Transcription activation domain GAL4.
DE Transcription factor; DNA binding domain;
KW Nuclear transport peptide; transcription activation domain; GAL4;
KW LexA; nuclear transport activity; transcription activation domain; ss.
KW nucleic acid transcription; ss.
OS Unidentified.
PN WO9849284-A1.
PD 05-NOV-1998.
PF 27-APR-1998; JP-309686.
PR 24-OCT-1997; JP-124795.
PR 28-APR-1997; JP-124795.
PA (HELI-) HELIX RES INST.
PI Noguchi T, Ueki N.
DR WPI; 99-024053/02.
DR P-PSDB; W85000.
PT Detection and isolation of DNA encoding peptides with nuclear transportability activity - by fusing candidate DNA with disabled transcription factor DNA and introducing into eukaryote with

promoter for transcription factor with reporter gene

Example 1: Pages 30-31: 62pp; Japanese.

The present sequence encodes transcription activation domain GAL4, which is used in the course of the invention. The specification describes a method for the identification of a DNA encoding a nuclear transport peptide. The method comprises fusing the DNA to DNA encoding a transcription factor which lacks the ability to transport a nuclear acid, and inserting the promoter region activated when the nuclear acid, having in its genome a promoter region of this promoter a transcription factor binds to it, and downstream of this promoter a reporter gene (such as LEU2 or beta-galactosidase). The transformed eukaryote is then cultured and if the reporter gene is expressed then the DNA tested encodes a peptide with nuclear transport activity. The disabled transcription factor is preferably a fusion protein containing an extranuclear signal sequence, a DNA binding domain (such as LexA) and a transcription activation domain (such as GAL4). The products can be used in the investigation and improvement of nucleic acid transcription, especially in connection with the development of new drugs.

Query Match 16.3%; Score 342; DB 1; Length 342;
Best Local Similarity 100.0%; Pred. No. 1.5e-82; Indels 0; Gaps 0;
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query 77 ggcgaatttaacaaagtggaataattgctgtagctcattgctcctcactactaac 136
DB 1 ggcgaatttttaacaaagtggaataattgctgtagctcattgctcctcactactaac 60
QY 137 agtagcaaggtccgaacctcaatacaactcaacaattctcaagcgtttcaaac 120
DB 61 agtagcaaggtccgaacctcaatacaactcaacaattctcaagcgtttcaaac 256
QY 197 attgctcctcactacgttcatactcaactcaataataataacacggttagtaaat 180
DB 121 attgctcctcactacgttcatactcaactcaataataataacacggttagtaaat 316
QY 257 gatgatgataataatcaaaacacctgacacctgttgagcagcaaacctggtat 240
DB 181 gatgatgataataatcaaaacacctgacacctgttgagcagcaaacctggtat 376
QY 317 cgcgttgatgataataatcaaaacacctgacacctgttgagcagcaaacctggtat 941
DB 241 cgcgttgatgataataatcaaaacacctgacacctgttgagcagcaaacctggtat 418
QY 377 ctatcgtatgataataatcaaaacacctgacacctgttgagcagcaaacctggtat 984
DB 301 ctatcgtatgataataatcaaaacacctgacacctgttgagcagcaaacctggtat

RESULT 6
ID 061607 standard: DNA; 2646 BP.
AC 061607:
DT 24-NOV-1994 (first entry)
DE Mutated GAL4 gene coding for Asp insertion between Asp863 and Val864.
KW Mutated GAL4; heterologous gene expression; enhanced; increased;
KW protein production; positive regulator; transcription activator;
KW galactose metabolism; ss.
OS Eukaryote.
FH Key
FT mat_peptide
FT Location/Qualifiers
FT 1..2646
FT /tag= a
FT /product= mutated GAL4
FT /transl_except= (pos:190..192, aa:lys)
FT /transl_except= (pos:757..759, aa:ser)
FT /transl_except= (pos:1377..1377, aa:glu)
FT /transl_except= (pos:1375..1377, aa:glu)
FT /transl_except= (pos:1990..1992, aa:cys)
PN J06078767-A.
PD 22-MAR-1994.
PF 07-SEP-1992; JP-262723.
PR 07-SEP-1992; JP-262723.
PA (GRBC) GREEN CROSS CORP.

THIS PAGE BLANK (USPTO)

Query Match	15.8%	Score 332.4	DB 1	Length 2646
Best Local Similarity	98.9%	Pred. No. 1.4e-79		
Matches 346		Mismatches 0	Indels 3	Gaps 1

RESULT	7
ID	Q61611 standard; DNA: 107 bp.
AC	Q61611;
DT	24-NOV-1994 (first entry)
DE	wild-type GAL4 C-terminal coding sequence.
KM	Mutated GAL4: heterologous gene expression; enhanced; increased;
KW	protein production; positive regulator; transcription activator;
KM	galactose metabolism; ds.
OS	Synthetic.
PH	Key
FT	misc_feature
FT	2..103
FT	/tag= a
FT	/note= "encodes amino acids 848-881 of wild-type
FT	GAL4"
FT	1..4
FT	/tag= b
FT	/label= sticky-end
FT	/note= "the complementary strand overhangs Q61611
FT	by the 4 base sequence 5'-TCGA-3'."
PN	J06078767-A.
PD	22-MAR-1994.
PF	07-SEP-1992; 262723.
PR	07-SEP-1992; JP-262723.
PA	(GREC) GREEN CROSS CORP.
WP	WIJ: 94-131274/16.
DR	P-PSDS: R62037.
PT	Mutated GAL4 and a method for the expression of foreign protein -
PT	useful for enhancing expression of protein
PT	Example 1: Page 14; 20pp; Japanese.
CC	A mutant GAL4 in which an Asp residue is inserted between wild-type
CC	Asp63 and Val864 is claimed. The mutated GAL4 can be used for

Query Match	4.9%	Score 103	DB 1	Length 107
Best Local Similarity	100.0%	Pred. No.	1	2e-18
Match 103, Conservative	0	Mismatches	0	Gaps 0

RESULT	8
ID	Q61613
AC	Q61613 standard; DNA; 110 BP.
DT	24-NOV-1994 (first entry)
DE	Mutated GAL4 C-terminal coding sequence.
KW	Mutated GAL4; heterologous gene expression; enhanced; increased;
KW	protein production; positive regulator; transcription activator;
KW	galactose metabolism; ds.
OS	Synthetic.
FH	Key
FT	misc_feature
FT	2..106
FT	/tag= a
FT	/note= "encodes amino acids 848-881 of wild-type
FT	GAL4 with additional Asp residue inserted
FT	between Asp 863 and Val 864"
FT	1..4
FT	/tag= b
FT	/label= sticky_end
FT	/note= "the complementary strand overhangs Q61613
FT	by the 4 base sequence 5'-TCGA-3'."

```

PN      J06078767-A.
PD      22-MAR-1994.
PE      07-SEP-1992; 262723.
PR      07-SEP-1992; JP-262723.
PA      (GREC ) GREEN CROSS CORP.
DR      WPI; 94-131274/16.
DQ      P-PSDB; R62038.
PT      Mutated GAL4 and a method for the expression of foreign protein -
PT      useful for enhancing expression of protein
PS      Example 1; Page 15; 20pp; Japanese.
CC      A mutant GAL4 in which an Asp residue is inserted between wild-type
CC      Asp863 and Val864 is claimed. The mutated GAL4 can be used for
CC      increasing the expression of foreign genes under the control of
CC      the GAL1, GAL7 or GAL10 promoters. Sequences coding for the
CC      C-terminal amino acids of wild-type and mutant GAL4 (Q61611 and
CC      Q61613, respectively) were each synthesised from two overlapping
CC      complementary oligonucleotides.
SQ      Sequence 110 BP; 42 A; 21 C; 20 G; 27 T;

```

Qy	316	CGCGTtTggaatcaactaagaaggatglttataaccacataagaag---gataatgtrataa	372
Db	1	CGCGTtTggaatcaactaagaaggatglttataaccacataagaagtrataa	60
Qy	373	ctatctatctgatgatgaagatacccccacccaacccccaaaaaagaag	418
Db	61	ctatctatctgatgatgaagatacccccacccaacccccaaaaaagaag	106

THIS PAGE BLANK (USPTO)


```
/product="isopropylmalate dehydrogenase"  
/protein_id="AAF20328.1"  
/db_xref="GI:6642854"  
/translation="MSAPKKIVYLRBDHVGQETTAIVKAIKISVRSNVKRFEPENH  
LIGGAIDITGVLPDEALEASKKVDVALLGAVGPKMGTSVRPOGLIKITKELOL  
YANLRPCNPNASDLIDLSPKIQFAGTGFVAVRVLGSIYFRRKRDGDGVAMDS  
QYTVPEVORITPRMAAFMALQHEPPLPISLIDANVYASSLWPKVEETIKNEPPLK  
VOHOLIDSAAMILVKNPRLNGIILTSNMGDIISDEASYIPGSLGLPSIASLSPD  
KNFARGIYPCGHSAPDLPRKNRVDPATITLISAMMLKSLINPEEKALIEDAVKRYLD  
AGIRGDLGSGSSTTEVGAVAEVKKIILA"  
5666..6375  
/note="ADCl promoter (PADH) cloned from PAS1; Durfee, T.  
1993"  
6383..76791  
/note="activation domain of GAL4 cloned from PGAD424"  
/codon_start=1  
/product="GAL4 DNA binding domain"  
/protein_id="AAF20329.1"  
/db_xref="GI:6642855"  
/translation="MDKAEILPEPPKRRKVELGTAAFNOSGNIADSSLSPTFNS  
NGPNLITQTNSOALSOPIASSNVHDMNNEITLASKIDGNSKPLSPGMDQTALN  
AGCITGMENTTMDVYNYRFDDEDTPPKKE"  
6797..6838  
/note="hemagglutinin epitope cloned from PAS1; Durfee, T.  
1993"  
6845..6899  
/note="multiple cloning site (NdeI, SfiI, NcoI, SmaI,  
BamHI, SalI, PstI) cloned from PAS1; Durfee, T. 1993"  
6917..7107  
/note="ADCl transcriptional terminator (TADH) cloned from  
PAS1; Durfee, T. 1993"  
BASE COUNT 2092 a 1637 c 1572 g 2103 t  
ORIGIN  
  
Query Match 16.5%; Score 346.6; DB 14; Length 7404;  
Best Local Similarity 96.2%; Pred. No. 4.6e-71;  
Matches 355; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
  
QY 50 aagaagaagcggaaggtcgtcgagcgcccaatttaacaaagtgggaatatgctgat 109  
|||||  
DB 6422 AAGAGAAAGGTGCAATGGGTACCGCGCAATTTTAATCAAGTGGGAATATGCTGAT 6481  
|||||  
QY 110 agctcatgtccttcaacttcaactacagtagcaagcggtccgaacctatacaactcaa 169  
|||||  
DB 6482 AGCTCATGTGCTTCACTTTCATACAGTAGCAACGGTCCGAACCTCATCAACACTCA 6541  
|||||  
QY 170 acaaatctcaagcgcttcaacaacaatgtgctctctcaagtcgatgataacttcgt 229  
|||||  
DB 6542 ACAAAATTCACACGCGCTTTCACAACCAATTCCTCTACGCTCATGATTAACCTTCATG 6601  
|||||  
QY 230 aataatgaatcaagcgtagtaagaaattgataatgaaataaacaacgtcacct 289  
|||||  
DB 6602 AATATATGAATACAGCGCTAGTAAATGATGATTAATTAATCAACCACTGTCACCT 6661  
|||||  
QY 290 ggttgacggaaccaactcgatataacgcggtttggaatcacatcacagggatgtttaacc 349  
|||||  
DB 6662 GGTTGACGAGCAACACTGCGTATACGCCGTTTGAATCCTACACGAGATGTTTAAATACC 6721  
|||||  
QY 350 actacaatgagatgatatataactatctatcgatgataagataccaccacaaaccca 409  
|||||  
DB 6722 ACTACAAATGATGATGATATATTAATCTATTCATGATGAAGATACCCCAACCAACCA 6781  
|||||  
QY 410 aaaaaaagag 418  
|||||  
DB 6782 AAAAAAGAG 6790  
  
RESULT 15  
A58525 7625 bp DNA Circular PAT 05-MAR-1998  
LOCUS A58525  
DEFINITION Sequence 29 from Patent WO9638555.  
ACCESSION A58525
```

```
VERSION A58525.1 GI:3714141  
KEYWORDS .  
SOURCE unidentified.  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 7625)  
AUTHORS Bogaert, T., Stringham, E. and Vandekerckhove, J.  
TITLE PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH CONTROL CELL  
BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND PHARMACEUTICAL COMPOSITIONS  
CONTAINING THEM AND THEIR USE IN THE CONTROL OF CELL BEHAVIOUR  
JOURNAL BOGAERT THIERRY (BE)  
COMMENT Patent: WO 9638555-A 29 05-DEC-1996;  
Other publication AU 6123496 961218.  
FEATURES  
source 1..7625  
Location/Qualifiers  
BASE COUNT 2201 a 1721 c 1573 g 2130 t  
ORIGIN  
  
Query Match 16.5%; Score 346.6; DB 5; Length 7625;  
Best Local Similarity 96.2%; Pred. No. 4.6e-71;  
Matches 355; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
  
QY 50 aagaagaagcggaaggtcgtcgagcgcccaatttaacaaagtgggaatatgctgat 109  
|||||  
DB 461 AAGAGAAAGGTGCAATGGGTACCGCGCAATTTTAATCAAGTGGGAATATGCTGAT 520  
|||||  
QY 110 agctcatgtccttcaacttcaactacagtagcaagcggtccgaacctatacaactcaa 169  
|||||  
DB 521 AGCTCATGTGCTTCACTTTCATACAGTAGCAACGGTCCGAACCTCATCAACACTCA 580  
|||||  
QY 170 acaaatctcaagcgcttcaacaacaatgtgctctctcaagtcgatgataacttcagt 229  
|||||  
DB 581 ACAAAATTCACAGCGCTTTCACACCAATTCCTCTCTAACGTCATATATCTTCAATG 640  
|||||  
QY 230 aataatgaatcaagcgtagtaanaattgataatgataatgataatgataatgataatg 289  
|||||  
DB 641 AATATATGAATACAGCGCTAGTAAATGATGATGATTAATTAATTAATTAATTAAT 700  
|||||  
QY 290 ggttgacggaaccaactcgatataacgcggtttggaatcacatcacacgggatgtttaacc 349  
|||||  
DB 701 GGTTGACGAGCAACACTGCGTATACCGCGTGGATGATCTACAGGATGTTTAAATACC 760  
|||||  
QY 350 actacaatgagatgatatataactatctatcgatgataagataccaccacaaaccca 409  
|||||  
DB 761 ACTACAAATGATGATGATATATTAATCTATTCATGATGAAGATACCCCAACCAACCA 820  
|||||  
QY 410 aaaaaaagag 418  
|||||  
DB 821 AAAAAAGAG 829  
  
Search completed: September 15, 2000, 10:01:53  
Job time: 2782 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2000, 09:15:31 ; Search time 1879.79 Seconds
(without alignments)
1993.766 Million cell updates/sec

Title: US-09-068-377-2
Perfect score: 2100
Sequence: 1 caatatctcaagctatacca.....gaaggaaaaaaaaaaaaaa 2100

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: GenEmbl:*
- 2: gb_da1:*
- 3: gb_da2:*
- 4: gb_da3:*
- 5: gb_da4:*
- 6: gb_da5:*
- 7: gb_da6:*
- 8: gb_da7:*
- 9: gb_da8:*
- 10: gb_da9:*
- 11: gb_da10:*
- 12: gb_da11:*
- 13: gb_da12:*
- 14: gb_da13:*
- 15: gb_da14:*
- 16: gb_da15:*
- 17: gb_da16:*
- 18: gb_da17:*
- 19: gb_da18:*
- 20: gb_da19:*
- 21: gb_da20:*
- 22: gb_da21:*
- 23: gb_da22:*
- 24: gb_da23:*
- 25: gb_da24:*
- 26: gb_da25:*
- 27: gb_da26:*
- 28: gb_da27:*
- 29: gb_da28:*
- 30: gb_da29:*
- 31: gb_da30:*
- 32: gb_da31:*
- 33: gb_da32:*
- 34: gb_da33:*
- 35: gb_da34:*
- 36: gb_da35:*
- 37: gb_da36:*
- 38: gb_da37:*
- 39: gb_da38:*
- 40: gb_da39:*
- 41: gb_da40:*
- 42: gb_da41:*
- 43: gb_da42:*

- 44: gb_da43:*
- 45: gb_da44:*
- 46: gb_da45:*
- 47: gb_da46:*
- 48: gb_da47:*
- 49: gb_da48:*
- 50: gb_da49:*
- 51: gb_da50:*
- 52: gb_da51:*
- 53: gb_da52:*
- 54: gb_da53:*
- 55: gb_da54:*
- 56: gb_da55:*
- 57: gb_da56:*
- 58: gb_da57:*
- 59: gb_da58:*
- 60: gb_da59:*
- 61: gb_da60:*
- 62: gb_da61:*
- 63: gb_da62:*
- 64: gb_da63:*
- 65: gb_da64:*
- 66: gb_da65:*
- 67: gb_da66:*
- 68: gb_da67:*
- 69: gb_da68:*
- 70: gb_da69:*
- 71: gb_da70:*
- 72: gb_da71:*
- 73: gb_da72:*
- 74: gb_da73:*
- 75: gb_da74:*
- 76: gb_da75:*
- 77: gb_da76:*
- 78: gb_da77:*
- 79: gb_da78:*
- 80: gb_da79:*
- 81: gb_da80:*
- 82: gb_da81:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1692	80.6	1692	MM087814	U87814 Mus musculu
2	1042.4	49.6	1656	HS094778	U94778 Human PEST
3	971.6	46.3	1428	AF038603	AF038603 Homo sapi
4	837.6	39.9	1382	AF038602	AF038602 Homo sapi
5	350.6	16.7	7653	AF033313	AF033313 Cloning v
6	350.2	16.7	7620	SV036450	U36450 E. coli/Yea
7	347.8	16.6	8117	CV029899	U29899 Cloning vec
8	346.8	16.5	6667	CV070024	U70024 Cloning vec
9	346.6	16.5	6650	CV031188	U31188 Cloning vec
10	346.6	16.5	6659	CV076447	U07647 Cloning vec
11	346.6	16.5	6665	CV070025	U70025 Cloning vec
12	346.6	16.5	6666	CV070026	U70026 Cloning vec
13	346.6	16.5	7389	AF134167	AF134167 Yeast 2-h
14	346.6	16.5	7404	AF134168	AF134168 Yeast 2-h
15	346.6	16.5	7625	AS58525	AS58525 Sequence 29
16	345.4	16.4	3271	SCYPR248C	Z73604 S.cerevisia
17	345.4	16.4	3694	YSCGAL4	K01486 Yeast (S.ce
18	345.4	16.4	11207	AB028139	AB028139 Drosophila
19	345.4	16.4	37808	SC38KXVI	Z67751 S.cerevisia
20	344.2	16.4	3028	AF149267	AF149267 Cloning v
21	343.8	16.4	980	AR037155	AR037155 Sequence
22	332.4	15.8	2645	E06922	E06922 DNA encodin
23	260	12.4	1005	MM018101	Y18101 Mus musculu
24	242.2	11.5	562	IR082015	AL10110 Homo sapi

C	45	61.6	2.9	32207	5	AR065852	
C	44	61.6	2.9	34189	82	KSU5064	
C	43	63.4	3.0	184113	81	KSU4958RAJ	
C	42	63.4	3.0	10850	14	U02455	
C	41	63.4	3.0	172281	81	EBV	
C	40	63.4	3.0	10737	14	XXU02428	
C	39	63.4	3.0	10596	5	I30503	
C	38	63.4	3.0	10506	5	I25041	
C	37	63.4	3.0	9600	5	A82665	
C	36	63.4	3.0	5852	14	U02454	
C	35	63.4	3.0	1150	81	HS4ULIR3	
C	34	66.4	3.2	180385	39	AC007461	
C	33	67	3.7	97348	12	AF091216	
C	32	77.2	3.7	258	13	G11561	
C	31	77.8	3.7	7218	5	I66494	
C	30	86	4.1	110	5	E06930	
C	29	90	4.3	110	5	E06929	
C	28	99	4.7	107	5	E06928	
C	27	103	4.9	107	5	E06927	
C	26	103	4.9	107	5	E06927	
C	25	214.2	10.2	476	12	MMDFE2	
C	24	214.2	10.2	476	12	MMDFE2	X6700 M.musculus
C	23	214.2	10.2	476	12	MMDFE2	E06927 Synthetic D
C	22	214.2	10.2	476	12	MMDFE2	E06928 Synthetic D
C	21	214.2	10.2	476	12	MMDFE2	E06929 Synthetic D
C	20	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	19	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	18	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	17	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	16	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	15	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	14	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	13	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	12	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	11	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	10	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	9	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	8	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	7	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	6	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	5	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	4	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	3	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	2	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	1	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	0	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-1	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-2	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-3	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-4	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-5	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-6	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-7	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-8	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-9	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-10	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-11	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-12	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-13	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-14	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-15	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-16	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-17	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-18	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-19	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-20	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-21	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-22	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-23	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-24	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-25	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-26	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-27	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-28	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-29	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-30	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-31	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-32	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-33	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-34	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-35	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-36	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-37	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-38	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-39	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-40	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-41	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-42	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-43	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-44	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D
C	-45	214.2	10.2	476	12	MMDFE2	E06930 Synthetic D

ALIGNMENTS

	Query Match	80.6%;	Score 1692;	DB 12;	Length 1692;	
	Best Local Similarity	100.0%;	Pred. No.0;			
	Matches 1692;	Conservative	100.0;	Mismatches 0;	Indels 0;	Gaps 0;
409	aaaaaagaggtaggtcgaccacagcgctcgcgtcttcctcatcttcgctgtatct					468
409	1 AAAAAAGAGGCTGGCTGACCCAGCGTCCGCTCTTCCTCATTTCCGCTGATTC					60
469	agccccaacaacaacaggtltgagccttcttcctcctccgagcttgacctgtgcttgg					528
469	61 AGCCCAACAACAACAGGTTGAGCCTTTTCCTCCTCCGACAGTTGCCTTGCTGTGG					120
529	ctgcctcttgagcgttcttcagaacggcgcgcgcttggagatltggaggaggagccttggactag					588
529	121 CTGCTCTTGTGAGCGTTTACAGCGGCGCGCTGGGATGGAGAGGAGGCTTGAGCTAGC					180
589	cgcgcttgtagctatggagactgtctctgtgctctcttgaccatgtcagaccctctctgaagcag					648
589	181 CGCGCTGGGACTGGGACCTGCTCTCTGAGCTCTGGCCATGCTCAGCCCTCTTGTAAGCAG					240
649	gagtgctcagcatlttgacacaagcgccctltgagatgtatgtggccagcttgcaattccagat					708

[illegible]

QY	769	ctlgagacgacagaagaatgtgcaagaatctggaagacgtctcagaacagaggccacagcg	828
Db	361	CTGAGACGGCGAGGAAGATGTGCAAGGATGTGGAGGAGCTGCTCAACACAGAGGGCCCAAGCG	420
QY	829	gagagagagtgacagaggaagagcctgtgtgacgaattgcacgcaaggtcgtgtgacagacagag	888
Db	421	GAGAGAGGATACGGGAGAGAGCTGTGTGCAGATTGCACCCAAAGGCTGTGTGGCGACAGAGAG	480
QY	889	atgaattccctctgagacctcctcttgactccctgaaacagcaacagaaatctggtgcagt	948
Db	481	ATGAATTCCCTTGAGGAGACTTCCTTGTGACTCCCTGAAGCAGCAAAACAGAAATGTGGCAGT	540
QY	949	gcacacatccagctgggccttggcccttggcgtggagagcttgggaagcttgggaagatttcga	1008
Db	541	GCACACATCCAGTGGCGCCCTGGGCCCTGCTGAGAGAGCTGCGAGCTGTGAGAGATTTCGA	600
QY	1009	gagagacagaagaagcagcgcgaaagaatctgagagccatcatgaaacctgtctcagaagaagc	1068
Db	601	GAGGACACAGAAAGCAGCGAGAGAAAGTATGAGGCCATCATGAGCCGTGCCAAGAGGC	660
QY	1069	aagttgttcgctctacaagaagaccatctgagttccaaagaagcatatgaccagaattgcag	1128
Db	661	AAGTTGTTCGCTCTACAAAGAACCATGAGGTCCAGAAAGCCATATGACCAAAAGTGCAGG	720
QY	1129	gattcagaatgattctgagcagccttcagagcgctgtgagtgtcaatctgacacagcaagaacaa	1188
Db	721	GATTCAGATGATTGCTGAGCGGCCCTTCGAGCGTGTGAGTCCAAATGGCCACCAGAGCAA	780
QY	1189	gtagaaaaagccacgaacaaagccaagcagctgcaagaagctcagccacagagcagaaga	1248
Db	781	GTAGAAAAGGCCGAGAACCAAGCCAAAGCCAGTGCAGAGAGTCAAGCCACAGAGGCGAAGAA	840
QY	1249	gtgtacagcagcaaatatctgaacaacctcggagagagagcagagaccgagttggagacaagaagac	1308
Db	841	GTTGTACAGCGCAAAATATTCGAAACAACCTGGAGAGAGCGCAGACCGAATGGAGAGAGAGC	900
QY	1309	cgagctacacctgtgaagaccttcacagtttcgaaggaattgacacggctcaccatcctccgaat	1368
Db	901	CGGACTACCTGTGAGGCCCTTCAGTTGTCAGAGATTTGACCGGCTACCCATCCTCCGCAAT	960
QY	1369	gccctgtgtgtgcactgttaacacagctctccatgtcagttgtgtcaagaagatgtagactcat	1428
Db	961	GCCCTGTGTGTGCACGTATACCAAGACTCTCCATCATCAAGTGTGTCAAGAGATGAGAGCTCTAT	1020
QY	1429	gaggaagtgtgcgtgcagaccttgagggtctgtgattgtgaaagtgtgaactcaatgatgtctaac	1488

|||||
 Db 1021 GAGGAAGTCGCGGTGACCTTGGGGCTGTGATGTGGAAGTACATCAATGAGCTTCATC 1080
 QY 1489 cagttccaaagaagactgagagagagcccaagctcaggtgctctatcaagaactactatgac 1548
 Db 1081 CAGTCCAAAGAGCACTGGCAGAGAGCCCGCCAGCTCCGGTGGCTTATCAAGAACTACTATGAC 1140
 QY 1549 agggaggtgaccacactgattggcagccctagatccagccctcctgagtgatgataag 1608
 Db 1141 AGGAGAGTACCCCATGATGTGGACGCCAGATCCAGCCCTCCGGTGTGATGAAG 1200
 QY 1609 aggttctcgtgagctgctacatggaaagcccaagacacacactcctcctgctgctcc 1668
 Db 1201 AGGTTCTCTGGGCTGTACATGGAAATGCCAAGACACACACTTCTGCTCTGCTTCC 1260
 QY 1669 aagagagactctactccccccttggaggaatgattgtgtcaagcatccatcgaagt 1728
 Db 1261 ACAGAGACTCTGCTCCACCCCTGAGCGGAATGAGTTGGTCTACGCATCCATCGAAGTG 1320
 QY 1729 caggcagaccagggaaaccttaactatcatcagccagagactaccgggactctacgactac 1788
 Db 1321 CAGGCGACCCAGGGAACCTTAATCATCAGCCAGGACTACCGGGGACATCTACGACTAC 1380
 QY 1789 actgcacagaattctgataagctgagacattccgcgggagacatcctgctgctatcctg 1848
 Db 1381 ACTGCACAGAAATCTGATGAGTGGAGCATTTCCGCGGAGACATCCCTGGGCTATCCTG 1440
 QY 1849 gaaggaggagatgagctgctgagactgtgagcggagaaacagatggtctgtgcccggg 1908
 Db 1441 GAAGGGAGAGATGGCTGGTGGAGCTGTGGAGCGGAGGAGCAAGCTGGCTTGTCCCTGGG 1500
 QY 1909 tctgacttgagaagctctgaggaagagctagagctctccacatactccgcctgactg 1968
 Db 1501 TCGTACTTGGAGAGCTCTGAGGAAGAGCTAGAGTCTCCACATACCTCCGCCCTGACTG 1560
 QY 1969 tgaagtcaggaactgttcttccatcacagccagagcctcagggcagaaaccaagccc 2028
 Db 1561 TGAAGTCAGGAGACTGTTCTTCCATCACGCCCTCAGGGGCGCAAGACCAACACCCC 1620
 QY 2029 ggttgagctgagcagatgagcgtggtgtgtgtgctgctcactcaataatgtgtcccaagagaaa 2088
 Db 1621 GGTGAGTGTGGGATGGCTGGGTGGCTGCTGCTGCTCAATTAATGTCTCCCAAGAGAAA 1680
 QY 2089 aaaaaaaaaa 2100
 Db 1681 AAAAAAAAAA 1692
 RESULT 2
 HSU94778 1656 bp mRNA PRI 05-JAN-1999
 LOCUS Human PEST phosphatase interacting protein homolog (H-PIP) mRNA,
 DEFINITION complete cds.
 ACCESSION U94778.1 GI:4100161
 VERSION 094778.1 GI:4100161
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 1656)
 AUTHORS Wilson,L.A., Fields,D., Cruz,L., Lasky,L., Friesen,J. and
 Simionovitch,K.A.
 TITLE The human homologue of mouse PIP-PIP interactor protein
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1656)
 AUTHORS Wilson,L.A., Fields,D., Cruz,L., Lasky,L., Friesen,J. and
 Simionovitch,K.A.
 TITLE Direct Submission
 JOURNAL Submitted (20-MAR-1997) Samuel Lunenfeld Research Institute, Mount
 Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5,
 Canada
 FEATURES Location/Qualifiers

source 1..1656
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_type="activated leukocyte"
 gene 217..1656
 /gene="H-PIP"
 CDS 217..1467
 /gene="H-PIP"
 /note="similar to Mus musculus PEST phosphatase
 interacting protein, encoded by GenBank Accession Number
 U87814, encodes potential N-terminal coiled coil domain"
 /codon_start=1
 /product="PEST phosphatase interacting protein homolog"
 /protein_id="A4D00762.1"
 /db_xref="gi:4100162"
 /translation="MMPQLQFKAFWRFDTAHTGVYEVLLQRLIDGRKMKDMRELLR
 QRHAHERYKELVQIRKRAAGGQTEINSLASPSLQOMENVSSTIQALTYREEL
 RSLKERERKREKRYEAVMDRVOKSLSLYRKAMSKTEHKDLSDDADAFER
 ISANGHIOKVEKQNKARQCKDSATEERYRQSLQLEVRARMEDEHRTCEAFOL
 QEDPRLTILNNALVHNSQLSMOCVKDEDEYEEVRLLEGSIDADIDSFQAKSTGT
 EPAPVRYONYVDREVPLTSSPQIOPSCMKRFSGLHSGPKTWSLAASASTETL
 TPTPERHEGYVTAIVQETIQGNPASPQERYRAFDYTAQNDELDSAGDILEVILEG
 EDGMWYVERNGQKGFVPGSYLEKL"
 misc_feature 430..490
 /gene="H-PIP"
 /note="encodes SH3 domain"
 polyA_signal 1600..1605
 /gene="H-PIP"
 BASE COUNT 384 a 472 c 537 g 263 t
 ORIGIN

Query Match 49.6%; Score 1042.4; DB 39; Length 1656;
 Best Local Similarity 80.0%; Pred. No. 1.6e-234;
 Matches 1304; Conservative 0; Mismatches 311; Indels 15; Gaps 6;
 QY 476 aacaaacaggttgagccttcttccctcccgagctgac-cctcgtgctgtgtgctcct 534
 Db 15 AACAACAGAGTGTGAGCTTTTCTCCCTCAGAGAGCTCTCTGCTCGGTGGCTGCT 74
 QY 535 tctgagcttcaagacggcgcgcgtgagtgagtgagagagagcctggtcagccgcgt 594
 Db 75 TCTGAGTGTTCAGACGGCGCGCGC-----CGGGAAGGGGGCGCTGGGCCACGCTTCC 128
 QY 595 gggactggaagc-tgctcctgcttggcccatctcagccctcctgaaagagagtg 653
 Db 129 AGGACTGGGAGGCTGCTGCTGCGCGCTGCCCTCCATAGGCCACACTGTGGCAGAGAG 188
 QY 654 ctgacatttgacaaacgccccttgagagatgagtgccagctgcaattccagatgacct 713
 Db 189 TGAGCTTTGCGCGGAGACGCCCTGAGAGATGATGCCCGACGTCAGTTCAAAAGATGCC 248
 QY 714 ctggtgcaaggaacttcaagcccaacagaggtatgaggtgctactcgaagagctgctga 773
 Db 249 TTGTTGACAGGAGCTTCACAGCCACACGCGCTACAGGCTGCTGTCGACAGCGCTTCTGA 308
 QY 774 cggcaggaagatgtgcaggaatgtgcagagctgctcagagagagggccagagcgagagga 833
 Db 309 TGGCAGAGAGATGTGCAAAAGACATGAGAGAGACTACTGAGGCAGAGGCCACAGCGAGGA 368
 QY 834 gaggtaacggaagagctgtgtcagatltcacgacagagctgtgtgacagagagatgaa 893
 Db 369 GCGGTACGGAGGAGAGCTGCTGAGATGCCACGGAAGAGAGTGGCCAGAGCAATCA 428
 QY 894 ttccctgaggaactccttgaactcctcgtgaagcagaacagagaatgtgtggaagtgcaca 953
 Db 429 CTCCTGAGGCGCTCTTGAAGCTCTTGAACAGCAAGATGAGATGTGGGAGCTCACA 488
 QY 954 catcaagctgagccctgagcccgctgagagagctgagagagcctgagagagcttcagagag 1013
 Db 489 CATCAGCTGAGCCCTGAGCCCTGAGAGAGCTGGGAGTGTCAAGAGATTCTCGGAGAG 548
 QY 1014 acagaaagagcagcgaagagatgagccatcatgagcctgttcagagagagcagtt 1073

```

Db      549 GCAGAGGACAGAGGAGGAGTATGAGGCCGTATGACCGGGTCCAGAGACCAAGCT 608
      1074 gtgcctctacaagaagaccatggaagtcacaagaagcatalgaccagaagtgcaaggaatgc 1133
      609 GTGCTCTTACAGAGAGCCCATGAGTCCAGAGACTTACAGACATTAAGTCCCTGGACTC 668
      1134 agaatagtctgaagcagccttgagcgtgtgagtgcacaaatggccacagaagcagaataga 1193
      669 GAGAGAGCGGATCACCCTTCGAGCCATTAGCGCAACGCCACCAAGAGAGGAGTGA 728
      1194 aaagagcagaacaagaagcagaagctgcaagagtgcaagcagaagcagaagaagatga 1253
      729 GAAGAGTCAGACAAACCCAGCGACGTCAAGAGACTCCGCCACCGACGACAGCGGGTATA 788
      1254 caagcaaaatalcgaacaacttggaagagcgaagaccgaagtgggaagcagaagcaggac 1313
      789 CAGGACAGACATTCGACGCTGGAAGAGTCCGCGCTGAGTGGAGCAGGACCGGAC 848
      1314 taactgtgaagccttccagtgtgcaagagtttgaccggtccatccctccgaatgacct 1373
      849 CACCTGTAGGCTTTGAGCTGCAAGAGTTTGACCGGCTGACATTCGCCGACAGCCCT 908
      1374 gtgggtgcaactgtaacagcctctcaatgcaatgcaatgcaatgcaatgcaatgcaatgca 1433
      909 GTGGGTGACAGACCAACCAAGCTCTCCATGCAAGTGTCAAGATGATGAGCTCTACAGAGA 968
      1434 agtgcagctgagcccttgaaggctgtgagtgtgaagtgatcaatgcaatgcaatgcaatgc 1493
      969 AGTCGGCTGACGCTGGAAGGCTGACAGCTGACGCCGACCCGACATCGCAATTCATCCAGGC 1028
      1494 caaagaagcagtgagagagagcccgagctccggtgctctatcagaactactatgacagga 1553
      1029 CAAAGACAGCGGACAGAGAGCCCGCTCCGCTGCTGCTTACCAAGAACTTACGATCGGGA 1088
      1554 ggtgaccccaactgtattgagcagcccttaagcagccctctgctgctgctgctgctgctgct 1613
      1089 GGTGACCCCGCTGACAGACAGCCCTGACATGACGCGTCTGCGGATGATGAAGAGGTT 1148
      1614 ctctgagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1670
      1149 CTCTGAGATCTGTGACGAGACCTCCCAAGACCACTTCTGTGGACCTTCTGTGCTGCTTCCAC 1208
      1671 agagaactgtgctccaccccttgagcaggaatgagtgctgtctacatcagatcgaagtga 1730
      1209 AGAGACCTTGACCCCGACCCCGAGGAGGAGATGAGGCTGTCTACAGCCATGCGACGTGA 1268
      1731 ggagaccagaagaacacttaactcaatcagccaggaactacagcggaactcagactaagac 1790
      1269 GGAGATTACAGGGAAMCCCGGCTTACAGCCAGGAGATVACCGGCGTTCTACGATTATATAC 1328
      1791 tgcacagaatctctgagctgaggaacttccgctgaggaagatccctgagcgatcattctga 1850
      1329 AGCGGAGAACCCAGATGAGCTGAGCTTCCCGGGAGACATCTCTGAGAGTATCTCTGA 1388
      1851 agggagagatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1910
      1389 AGGGAGAGATGCTGCTGTGAGCTGTGAGAGGAAAGGAGCGGTGGTGTCTGCTGCTGCTGCT 1448
      1911 gtacttgaagaagctctgaggaagcctagcaatctccacatcctccgctgctgctgctgctgctgct 1970
      1449 CTACTGTGAGAGCTTTTGTGAGAGGCGCCAGGAGCCCTTTCGAGACTT - - - - - 1506
      1971 aggttaagagactgttcttctcctacacgagcagagctcagcgagcagaagacagcgag 2030
      1507 GAGCCAGAGAGTCCCGCCAGACTGTC - - - - - 1564
      2031 tgggtctgagcagtgagcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 2090
      1565 CCGAGCCCGGAGAGAGGAGCGCTGTCTGCTCCAGGGAATTAAGAGAGTGGCTTCTGTTAAAA 1624
      2091 aaaaaaaaaa 2100
      1111111111

```

```

Db      1625 AAAAAAAAAA 1634

RESULT      3
AF038603    1428 bp      mRNA      PRI      02-FEB-1999
LOCUS      AF038603
DEFINITION Homo sapiens CD2 binding protein 1 long form mRNA, complete cds.
ACCESSION AF038603
VERSION    AF038603.1
KEYWORDS   GI:2921550
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 1428)
            Li,J., Nishizawa,K., An,W., Hussey,R.E., Liallos,F.E., Salgia,R.,
            Sunder-Plassmann,R. and Reinherz,E.L.
            A cd15-like adaptor protein (CD2BP1) interacts with the CD2
            cytoplasmic domain and regulates CD2-triggered adhesion
            EMBO J. 17 (24), 7320-7336 (1998)
JOURNAL    2 (bases 1 to 1428)
MEDLINE    99077800
REFERENCE  2 (bases 1 to 1428)
AUTHORS    Li,J., Nishizawa,K., An,W., Hussey,R.E., Liallos,F.E.,
            Sunder-Plassmann,R. and Reinherz,E.L.
            Direct Submission
            Submitted (15-DEC-1997) Immunobiology, Dana-Farber Cancer
            Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES   Location/Qualifiers
            source          1..1428
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /cell_type="T cell"
                        1..1251
                        /note="CD2BP1"
                        /codon_start=1
                        /product="CD2 binding protein 1 long form"
                        /protein_id="A01195.1"
                        /db_xref="GI:2921551"
                        /translation="MMPQLQPKDAFWCRDFTAHNGYEVLLQRLDGRKCKDMELLR
                        ORAOEERYGKELVQIARKAGQTEINLSRASFSLKQOMENVSSTHQLATLREEL
                        RSLFEPRERKORRYEAVMDRQKSLSTYKAMESEKTYEGCRPADAEQAFER
                        ISANGHQKQVRSKONKARCKDSDNTEARVROSIAOLEKXRAEMEOHRHTCEAFOL
                        QEPDRLLTLRNALWHSNQLSMQCYKDELEIYRATLEEGSIDADISFIOASTGT
                        EEPAPVPQNYDYREVPPLTSSPIQSPCGMIKRPSTGLHSPKTSIAASNASTETL
                        TPTPERNGEYVTAIAVDIIGNPASPAPQETRYALDYDTAONPDELDLSAGDILEVILEG
                        EDGWMTEVRNGORGFVPSYLEKL"
BASE COUNT 352 a      396 c      462 g      218 t
ORIGIN
Query Match      46.3%; Score 971.6; DB 39; Length 1428;
Best Local Similarity 81.6%; Pred. No. 7.1e-218;
Matches 1161; Conservative 0; Mismatches 254; Indels 7; Gaps 3;

```

OY	982	gagctgcyggaagccttggagaagltccgagagagacagaagaagcaagcggaagaatgatgag	1041
Db	301	GAGCTGCGGAGTCTCCAGAGATTTCGTGAAGGACAGAAAGACGAGAGAAATATGAG	360
OY	1042	gcatcatgacccgtgttccagaagaagaattgtgcctccagaagaagacatagagttcc	1101
Db	361	GCCGTCATGAGCCGGGTCCAGAAAGACACACTGCTCTACAAAGAGGCCATGAGTCC	420
OY	1102	aagaaggaatgatgaccagaagaagtgcaggaatgacagatgactgaagcccttcagcgt	1161
Db	421	AAGAAGACATATGAGAGCAGAAAGTCCGGGACGCGGACACGCGGAGCGAGGCTTGACAGCG	480
OY	1162	gtgaaatgttgcgaatgagccaccagaagaagttagaaaaagagccagaagaagccaagctg	1221
Db	481	ATTAGCGCCACGCGCCACAGAAAGCAGGTGAGAAAGATCAGAAACCAAGCCAGCAGTGC	540
OY	1222	aaggaatgcacccaagaagagcagaagaagtgtacagggccaataatcgaacaactgtgagaga	1281
Db	541	AAGGACTCGGCCAACCGAGGAGAGCGGGTATACAGGACAGAGCAATTGGCAGCTGGAGAG	600
OY	1282	gcgaggaagccagattggagacagagagcaaccggaactaactgtgaggcttcagttgcagag	1341
Db	601	GTCCGGGCTGAGTGGGAGAGAGAGACACCGGACCACTGTAGGGCTTTACAGCTGCAAGAG	660
OY	1342	tttgacaggtcacaactccctccgcaatgccccttggatggactgttaacagactccatg	1401
Db	661	TTTTGACCGGCTGACCTTTCTCCGCAACGCCCTGTGGGTGCACAGCAACCACTTCCATG	720
OY	1402	cagtgatgacaagatgatgagctctatlgaaagaatgcgctgtgaaccttgaggctgtagt	1461
Db	721	CAGTGTGTCAAAGATATGATAGCTCTCTCAGAGGAATGCGCTGACGCTGGAAGAGCTCAGC	780
OY	1462	gtggaagtgaaatcaatgatgcttcatccaggtccaaagaagactgtgagaagccccagct	1521
Db	781	ATAGACCGCCGACATGCACAGATTTCATCCAGGCCAAGAGCACGGGCACAGAGCCCCGCT	840
OY	1522	ccggtgcttatacagaactactatgacaggaagtgaaacccactgatatgagccctcag	1581
Db	841	CCGGTCCCTACACAGAACTATTACGATCGGAGGAGGTCACCCCGCTGACAGAGCCCTGGC	900
OY	1582	atccagacctctgcggtgtgataaagaagttctctggctgtctacatlgaaatgccaa	1641
Db	901	ATACACCCCTCTCGCGGCATGTAAAGAGGTTCTTGGAATGTGTGCACGGAAAGTCCAG	960
OY	1642	accacaacct---ctgctctctgtgtcttccagaagaacttgactccaccacctgaagc	1698
Db	961	ACCACTTCGTGGACACTTCTGTGTGCTCCACAGAGACCTTGACCCCCACCCCGAAGCG	1020
OY	1699	aatgattgtgtcagcatccalcagaaatgacagtgcaagcgaaaccaagggaaaacttaactca	1758
Db	1021	AATGAGGGGTCTACACACGCCAATCGCAGTGCAGAGATACAGGAAACCCGGCTCACCA	1080
OY	1759	gccccagactaccgggacctctcagactacacatgcacaagaatctgtagtgctgagacat	1818
Db	1081	GCCCGAGATACCGGGCGCTCTACGATTTATACAGCCGAGAACCCAGATGAGCTGACCTG	1140
OY	1819	tccgcgaggaagacatccctgcygtcatctctgaaagggagagatgctgtgtgaaatgtgag	1878
Db	1141	TCCGGCGGAGACATCCTGGAGGTGATCTCTGGAAAGGGAGGATGCTGTGAGACTGTGAG	1200
OY	1879	cggaaacggacaacgtgtgcttltgtccctgggtgcgtacttggagaagctctgagaagaagct	1938
Db	1201	AGGAACGGGACAGCTGCTTCGTCCCTGCTCTACTCTGGAGAAACCTTTGAGGAAAGGCC	1260
OY	1939	agcagttctcacatlaactccgcgcttactctgaggttgagagcaagactgtttcttcaacag	1998
Db	1261	AGGAGCCCTTGGAGCT--GCCCTGCAATGAGGACGACAGAGTGCCTCCACGACACTGTC--	1316
OY	1999	cccagagctcaaggggcagaaaccaagcccgtgtgtgtgtgagcatggagtgtgtgtcgc	2058
Db	1317	CCCACTTCTATGCGGCTCAGAACCAAGCGTCCCCACGCCCGGAGAGGAGACCTGTCTCT	1376

OY	2055	tactctcataaatgtcttcccgaggaagaaaaaa	2100
Db	1377	CCGAGGAATAAAGAGTCGTTCTTCTTA	1418
RESULT	4		
AF038602			
LOCUS	AF038602	1382 bp	mRNA PRI 02-FEB-1999
DEFINITION	Homo sapiens CD2 binding protein 1 short form mRNA, complete cds.		
ACCESSION	AF038602		
VERSION	AF038602.1	GI:2921548	
KEYWORDS	.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	(bases 1 to 1382) Li,J., Nishizawa,K., An,W., Hussey,R.E., Liaillos,F.E., Salgia,R., Sunder-Plassmann,R. and Reinherz,E.L. A cdcl5-like adaptor protein (CD2BPI) interacts with the CD2 cytoplasmic domain and regulates CD2-triggered adhesion EMBO J. 17 (24), 7320-7336 (1998)		
TITLE	2 (bases 1 to 1382) Li,J., Nishizawa,K., An,W., Hussey,R.E., Liaillos,F.E., Sunder-Plassmann,R. and Reinherz,E.L. Direct Submission Submitted (15-DEC-1997) Immunobiology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA		
JOURNAL	Location/Qualifiers		
MEDLINE	1..1382		
AUTHORS	/organism="Homo sapiens"		
FEATURES	/db_xref="taxon:9606"		
SOURCE	/cell_type="T cell"		
CDS	1..1194 /note="CD2BP1s" /codon_start=1 /product="CD2 binding protein 1 short form" /protein_id="AAD11958.1" /db_xref="GI:2921549" /translation="MMPOLQFKDAFMNCFDPAHTGYEVLRLDLGRMCKDMELFL ORAAEREGKELVOLAKRAGSGTETINSLSFSLSKOMENVSSTQLVLTKEEI RSLEEFREQRKRKYEAVDVRQKSLSYLKAMESEKTYEDCRADADEAFOL ISANGHOKVERKSONKAOCACDYAEARVROGIAOLEKRAEDEDHRTCAFOI OEPRFLTIRNALWHSNOLSMOVCXKDELEEVRLITEGSGIDADIDSFIQAISTGT EPPEVRIADSAASRFGLSGSPKTSLAASASTPTLPPTPERNBVTYAIVQEII OGNPASAPQETRALCYDTAONPDELID.SAGDILEVTLLEGEDGMFTVERNGRGVPSS YLEKL"		
BASE COUNT	344 a . 375 c 453 g 210 t		
ORIGIN			
Query Match	39.9%; Score 837.6; DB 39; Length 1382;		
Best Local Similarity	79.4%; Pred. No. 2,1e-186;		
Matches 1070; Conservative	0; Mismatches 214; Indels 64; Gaps 4;		
OY	682	atgatagcacacagtcaattcgcgaagatgccttcgttgtagaggaccttaacggcccacaca	741
Db	1	ATGATGCCCCCACTCACAAGTCAAAGATGCCTTTTGATGCGAGACTTTCACACGCCACAG	60
OY	742	gagla tgagtg tctact ctacaga agtgtct gtagac ggcagacaga aaatg tgc aaagat gtgag	801
Db	61	GCGTACGAGGTCCTCTCGACGGCGCTTTCGATGGCAGGAAGATGTGGCAAGAACATGAG	120
OY	802	gagctgtctacagacagagggccagcgcgaggaagaggtacaggaaagagctggtcacagt	861
Db	121	GAGCTACTGAGGCAAGAGGCCCCAGGCGGAGGAGCGGTACGGGGAAGAGAGCTGGTCAGATC	180
OY	862	gaagcgaagcgtggtgtggcgacagacagatatgaattccctcgagaacctctttaccctctg	921
Db	181	GCACGGAAGCGCAGGTGCGCCACAGACGAGATCAACTCCCTGAGAGGCGCTCTTGTACTCTTG	240
OY	922	aagcagcaaacagagatgtggtgacgtgcacacatccacgtctgagccctgagccctgtgag	981

```

Db 241 AAGCAGCAAAATGAGATTTGGGAGCTCACACATCCAGCTGGCGCTGACCTCGTGGAG 300
Oy 982 gagctgcgagccttgagagattccgagagagacagaagaagcagcgagaaagatagag 1041
Db 301 GAGCTGGGAGATCTCGAGGAGATTCTGTAGAGCAGAGAGCGAGAGAGAGAGATATGAG 360
Oy 1042 gccatcatgagccgtgtccaggaagaagattgtcgtcttacaagaagaagcattgagtc 1101
Db 361 GCGGTCAATGAGACCGGTCACAAAGAGCAAGCTGTGCTCTACAAAGAGCGCATGAGTCC 420
Oy 1102 aagaagcatalagaccagaagttgagagatgcatatgtctgagcagccttcgagcgt 1161
Db 421 AAGAAAGCAATACAGAGCAGAGAGTGGCGGAGCGGAGCGGAGAGCGGCTTGAGAGCG 480
Oy 1162 gtgagtgccaatgagccagcagaagcagaatgaaagaagccagaacaaagccagatgac 1221
Db 481 ATTAGCGCCCAAGCGCCACCAAGAACAGGTGAGAAAGTCAAGCAAGCCAGCGAGTGC 540
Oy 1222 aaggaagtcacacacagagcgagaaagagttgacagggcaaatatcgaaacactgagaga 1281
Db 541 AAGGACTTCGGCCACCGAGGAGAGCGGCTATACAGGCGAGACATTTCCGCTGAGAGAG 600
Oy 1282 gcgagagcagagtgagagcagagcagcagcagcagcagcagcagcagcagcagcagcag 1341
Db 601 GTCGGGCTGAGTGGAGAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 660
Oy 1342 ttgagcagcgtcacatcctcgcagagcagcagcagcagcagcagcagcagcagcagcag 1401
Db 661 TTATACCGCGCTGACATTTCTCCGCAACGCTGTGGTGACAGCAACAGCTGTCATG 720
Oy 1402 cagtggttcaagagatgagcctatgagaaatgacgagcagcagcagcagcagcagcagcag 1461
Db 721 CAGTGTCTCAAGAGATGATGAGCTCTACAGAGAAAGTCCGCTGAGCCTGGAAGCTGCAGC 780
Oy 1462 gtgagaggtgacatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1521
Db 781 ATAGAGCGCGCATGACAGATTTTCATCCAGGCAAGAGCAGCGGAGAGAGCGCCCGCTG 840
Oy 1522 ccggtgtccttccgaactactatgacagagagtgagcccaatgtgagcagcctagc 1581
Db 841 GAGG----- 844
Oy 1582 atccagcctcctcgtggtgataaagaggttctctgagcgtgctacatgagagtcctcaag 1641
Db 845 -TCCGGCTTGGGAGAGGAGCGCTCTAGGTTCTCTGAGCTGCTCAGAGAGTCCCAAG 903
Oy 1642 accaaccct---ctgctcctggtgtcctcagagagcagcagcagcagcagcagcagcagcag 1698
Db 904 ACCACTTCGTGGGAGGCTTGTGCTGCTGCACAGAGAGACCTGACCCCAACCCCGAGCGG 963
Oy 1699 aatgagttggtctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1758
Db 964 AATGAGGCTGTCTACAGCCATTCAGAGTGCAGAGATACAGAGAAACCCGGCTTCACCA 1023
Oy 1759 gccagagactaccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1818
Db 1024 GCCCAGAGATACCGGGCGCTCTACGATTATACAGCGCAGAACCCAGATGAGCTGAGACTG 1083
Oy 1819 tccgagagagacatcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1878
Db 1084 TCCGGGGAGACATCTGAGAGTGTCTGCTGAAAGGAGAGGATGAGTGTGAGTGTGAGG 1143
Oy 1879 cggagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1938
Db 1144 AGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1203
Oy 1939 agcagttccacacatcctcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1998
Db 1204 AGGAGCGCTTCTGAGACT--GCCCTGCCAGTGTGAGGAGCAGACATGCCCCAGCAGCTGTC-- 1259
Oy 1999 cccaggtcctacaggggccaagaaccaagc 2026

```

```

Db 1260 CCCACTTGTAGGGCCCGAAGCAAGC 1287
RESULT 5
AF033313 7653 bp DNA circular SYN 25-JAN-1999
LOCUS AF033313
DEFINITION Cloning vector PADgal4 2.1, complete sequence.
ACCESSION AF033313
VERSION AF033313.1 GI:4185574
KEYWORDS Cloning vector PADgal4 2.1.
SOURCE Cloning vector PADgal4 2.1.
ORGANISM Cloning vector PADgal4 2.1.
REFERENCE Cloning vector PADgal4 2.1.
AUTHORS artificial sequence; vectors.
TITLE 1 (bases 1 to 7653)
JOURNML Direct Submission
AUTHORS Submitted (06-NOV-1997) Technical Services, Stratagene, 11011 North
Torrey Pines Road, La Jolla, CA 92037, USA
REFERENCE 2 (bases 1 to 7653).
AUTHORS Lu, Q.
TITLE Direct Submission
JOURNML Submitted (25-JAN-1999) Technical Services, Stratagene, 11011 North
Torrey Pines Road, La Jolla, CA 92037, USA
REMARK Sequence update by submitter
COMMENT On Jan 25, 1999 this sequence version replaced gi:2695664.
FEATURES
source
location/Qualifiers
1..7653
/organism="Cloning vector PADgal4 2.1"
/specific_host="Saccharomyces cerevisiae"
/db_xref="taxon:69544"
/note="yeast two-hybrid plasmid vector"
BASE COUNT 2184 a 1738 c 1607 g 2124 t
ORIGIN

```

```

Query Match 16.7% Score 350.6; DB 14; Length 7653;
Best Local Similarity 89.5%; Pred. No. 5.3e-72;
Matches 377; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

```

```

Oy 50 aagaagaagcggaaggtctcgcagcgcgccaaatttaacaaagtggaataatgtcgtat 109
Db 461 AAGAGAAAGGTGAGATTTGGGTACCGCGCCATTTTAAATCAAAAGTGGAAATTTGCTGAT 520
Oy 110 agctcattgtccttcaacttcaactaagtagaagcagcagcagcagcagcagcagcagcagcag 169
Db 521 ACCTCATTTGCTTCACTTCACTTAACAGTACCAAGCGTCCGACCTCATTAACACTCA 580
Oy 170 acaaatctcaagcgcttcacacacaaattgctcctcctcaacgcttcatataactcag 229
Db 581 ACMAATTCANAGCGCTTTCACACCAATGGCTCTCTAAGCTTCATATAACTCATG 640
Oy 230 aataatgaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 289
Db 641 AATATGAAATACACGGCTAGTAAATGTGATGATGATGATGATGATGATGATGATGATGATGAT 700
Oy 290 ggttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 349
Db 701 GGTTCAGCGACCAACCTCGTATTAACGGCTTGGATGATGATGATGATGATGATGATGATGATGAT 760
Oy 350 actacatgagatgatatataactatcctatcagatgagatgagatgagatgagatgagatgagatgag 409
Db 761 ACTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 820
Oy 410 aaaaaagagtggtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 469
Db 821 AAAAAAGATGAAATTAAGATTCCTGCTAGAGAGAAATTAATCTCTAATGCTTCTC 880
Oy 470 g 470
Db 881 G 881
RESULT 6

```


SVU36450
LOCUS SVU36450 7620 bp DNA circular SYN 02-NOV-1995
DEFINITION E. coli/Yeast shuttle vector pAD-GAL4, complete sequence.
ACCESSION U36450
VERSION U36450.1 GI:1049256
KEYWORDS
SOURCE Shuttle vector pAD-GAL4.
ORGANISM Shuttle vector pAD-GAL4.
REFERENCE 1 (bases 1 to 7620)
AUTHORS Mullinax,B.
TITLE Hybridap Two-Hybrid Vector System for Protein-Protein Interactions
JOURNAL Strateg. Mol. Biol. 8, 3-5 (1995)
AUTHORS Mullinax,B.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1995) Becky Mullinax, Research and Development,
Stratagene, 11011 North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
source
1. .7620
/organism="Shuttle vector pAD-GAL4"
/plasmid="pAD-GAL4"
/db_xref="taxon:43132"
/note="composite sequence, has not been sequenced"
/gene="LEU2"
complement(1582. .2676)
/gene="LEU2"
/codon_start=1
/transl_table=11
/product="3-Isopropylmalate dehydrogenase"
/protein_id="AA80348.1"
/db_xref="GI:1049257"
/translation="MSAPKRTIVLPDGHVGOETIAEAIKYLKAIIDVRSNVPKDFENH
LUGAIDATGAVLPDLEALSAKGVAVILGAVGPMKGVSRPEGLIKIRKELQ
YANLRPCNRASDLSLDLSPKPKAKTDVYVRELVGITFEGRKEDDGCYAMDS
QYVPEVORTIRMAAFMALQHEPPLPWSLDKANVLASSLMKRYEETIKNEPTLK
VOHQLIDSAMITVKNPHTLNGIITSNMGDITSDSAVIPSIGLIPSAIASLPD
KNRFGILYERCHGSAADLPKKNVPIATITLSAAMMLKLSINLPEEGKAIDAVKVVAD
AGIRTDLGGSNSTTEVGDVAEVEKKILA"
complement(5209. .6069)
CDS
/codon_start=1
/transl_table=11
/product="beta-lactamase"
/protein_id="AA80349.1"
/db_xref="GI:1049258"
/translation="MSIQHRVALIIPFAFCIPVPAHPETLVKVKDAEDQLGARVY
IELDLNSGKILDSFRPEERPMKSTFKVLGCVLSRIDGQDQLGRIRHNSDYVE
YSPVTEKHLIDGMTVELCSAALTMSDNTANLLLTIGPKBELTALHNMGSQVPL
DRWPELNEAIIPNDERDTPVAMATTLIRKLIGELLTSLRSQOLIDMPEADKVAQPL
LRNALPAGWFIADKSGAGERSGIIAALGPDGKPSRIIVITTTGSGATMDERNROIA
EIGASLIKHW"
BASE COUNT 2117 a, 1738 c 1605 g 2160 t
ORIGIN
Query Match 16.7%; Score 350.2; DB 14; Length 7620;
Best Local Similarity 92.9%; Pred. No. 6,6e-72;
Matches 367; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Db 641 AATAATGAATACACGGCTAGTAAATGTGATGATTAATTCAAACCACTGTCACT
Oy 290 ggttgacggacccaactgcgcataaccgcttggatcaccacagagatgtttatcc 349
Db 701 GGTGACGGACCAACTGCGTATACCGGTTGGAACTACACGAGGATGTTTATACCC 760
Oy 350 actacaatgatalgatalataactatcatalcatalgatalgaatataccaccacaacca 409
Db 761 ACTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 820
Oy 410 aaaaaagggtgggtgcgaccacgcgtccgc 444
Db 821 AAAAAAGATGCAATTCGCCGCCCTCGAGCC 855
RESULT 7
LOCUS CVU29899/c 8117 bp DNA circular SYN 08-JAN-1997
DEFINITION Cloning vector pACr2 MatchmakerII, complete sequence.
ACCESSION U29899
VERSION U29899.1 GI:1262182
KEYWORDS
SOURCE Cloning vector pACr2.
ORGANISM Cloning vector pACr2.
REFERENCE 1 (bases 1 to 8117)
AUTHORS Kitz,P.A.
TITLE Clontech Vectors on Disk, version 1.3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 8117)
AUTHORS Stille,J.S.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-1995) John S. Stille, Clontech Laboratories, Inc.,
4030 Fabian Way, Palo Alto, CA 94303, USA
COMMENT
This vector can be obtained from Clontech Laboratories, Inc., 4030
Fabian Way, Palo Alto, CA 94303, USA. To place an order call (415)
424-8222 or (800) 662-2566, extension 1. International customers,
please contact your local distributor. For technical information,
call (415) 424-8222 or (800) 662-2566, extension 3. This sequence
has been compiled from information in the sequence databases,
published literature and other sources, together with partial
sequences obtained by Clontech. If you suspect there is an error
in this sequence, please contact Clontech's Technical Service
Department at (415) 424-8222 or (800) 662-2566, extension 3 or
E-mail TECH@CLONTECH.COM.
FEATURES
source
1. .8117
/organism="Cloning vector pACr2"
/db_xref="taxon:41435"
1. .2057
/note="Yeast 2 micron ori"
2474. .3568
CDS
rep-origln
/codon_start=1
/transl_table=11
/product="LEU2"
/protein_id="AA839920.1"
/db_xref="GI:915410"
/translation="MSAPKRTIVLPDGHVGOETIAEAIKYLKAIIDVRSNVPKDFENH
LUGAIDATGAVLPDLEALSAKGVAVILGAVGPMKGVSRPEGLIKIRKELQ
YANLRPCNRASDLSLDLSPKPKAKTDVYVRELVGITFEGRKEDDGCYAMDS
QYVPEVORTIRMAAFMALQHEPPLPWSLDKANVLASSLMKRYEETIKNEPTLK
VOHQLIDSAMITVKNPHTLNGIITSNMGDITSDSAVIPSIGLIPSAIASLPD
KNRFGILYERCHGSAADLPKKNVPIATITLSAAMMLKLSINLPEEGKAIDAVKVVAD
AGIRTDLGGSNSTTEVGDVAEVEKKILA"
4268. .4327
/note="lox Site"
4367. .4412
/note="lox Site"
4415. .4742
misc_feature
terminator
misc_feature
5042. .5068

CDS
 /note="encodes HA epitope which fuses to Gal 4"
 complement(5081..5488)
 /note="encodes 5' region of the Gal 4 activation domain
 polypeptide: stop codon is located downstream within the
 multiple cloning site"
 /codon_start=1
 /transl_table=11
 /product="Gal 4 activation domain polypeptide"
 /protein_id="AB39921.1"
 /db_xref="GI:915411"
 /translation="MDKALIPPEPKKKRYEIGTAAFNOSGNITADSSISTFTNSS
 NGPNLITTTQTSQALSDPIASSNADNPNNEITRASKIDDDGNNKPLSPGWTDTAYN
 AFGITGMENTTDDVYNYLFDDDETPNPKE"
 5504..5901
 /note="S. cerevisiae ADH1 Promoter"
 6336..6979
 /note="pBR322 origin of replication"
 complement(7127..7987)
 /codon_start=1
 /transl_table=11
 /product="beta-lactamase"
 /protein_id="AB39922.1"
 /db_xref="GI:915412"
 /translation="MSIOHFRVALIPFAFLPYFAHPETLVKVKDAEDQLGARVY
 IELDNGSKILESFRPEERPMWSTFKVLLCGAVISRDAGQELGRIRHSQNDLVE
 YSPVTEKHITDGMVRELCSAIIITMSDNTAAILTTIGSPKEFLIPLNMGDHVTRL
 DRWEPELNAIPNDERDITTPAAMATTLRLKLLIGELLTASRQQLIDWMEADKVAGPL
 LRSALPAGMFIADKSGAGERSRGIIAALGPGKPSRIYVITTSQATMDERNNOIA
 EIGASLIKHW"
 8029..8034
 -35-signal
 8052..8057
 BASE COUNT 2424 a 1676 c 1757 g 2260 t
 ORIGIN

Query Match 16.6%; Score 347.8; DB 14; Length 8117;
 Best Local Similarity 93.1%; Pred. No. 2.4e-71;
 Matches 364; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 50 aagaagaacggaaggtcgcagagcgagcccaatttaacaaagtggaattgctgat 109
 |||||
 DB 5449 AAGAGAAAGCTGGAATGGGTACCGCCCAATTTTAATCAAAAGTGGCAATATTGCTGAT 5390
 QY 110 agctaatgcttcacttcacttaacagtagcaacgctcgaacctcataacaaactcaa 169
 |||||
 DB 5389 AACTCATTTGCTTCACTTACTACAGTAGCAACGGTCCGAACCTCATTAACAACACTCAA 5330
 QY 170 acaaatctcaagcgttccaacaaatgctcctcctcaagttcatgataactcatg 229
 |||||
 DB 5329 ACAAATTCACAAGCGCTTACAAACCAATTCCTCTTAACGTTGATGATTAACCTTCATG 5270
 QY 230 aataatgaatacagcgtagtaaatgtagtgaatgaatcaaatcaaacacgtcaact 289
 |||||
 DB 5265 AATTAATGAATACAGCGCTTATGTAATTAATGATGATTAATTAATTAATTAATTAATTAAT 5210
 QY 290 ggttggacgagcaaacactgcgtataaagcgtttggaatacactcaaggaatggttaatacc 349
 |||||
 DB 5209 GGTGGACGAGCAAACTGCGTATACGGGTTGAATFACCTACAGGATGTTTAATACC 5150
 QY 350 actcaaatgtagatgtatataactctctatcgtatgtagaagatacccccaaaccca 409
 |||||
 DB 5149 ACTACATGATGATCATATATATATATCTATCTATTCATGATGATGAAGATACCCACCAACCA 5090
 QY 410 aaaaaaaggagtggtgcgaacccacgcgtccg 440
 |||||
 DB 5089 AAAAAAGAGATCTGTATGGCTTACCATACG 5059

RESULT 8
 CUV70024 6667 bp DNA SYN 21-MAR-1997
 LOCUS CUV70024
 DEFINITION Cloning vector pCAD-C1 complete sequence.
 ACCESSION U70024

VERSION U70024.1 GI:1595839
 KEYWORDS
 SOURCE Cloning vector pCAD-C1.
 ORGANISM Cloning vector pCAD-C1
 artificial sequence; vectors.
 REFERENCE
 1 (bases 1 to 6667)
 James,P., Halladay,J. and Craig,E.A.
 Genomic Libraries and a host strain designed for highly efficient
 two-hybrid selection in yeast
 Genetics 144 (4), 1425-1436 (1996)
 JOURNAL MEDLINE
 97132579
 REFERENCE
 2 (bases 1 to 6667)
 James,P., Halladay,J. and Craig,E.A.
 Direct Submision
 Submitted (09-SEP-1996) Biomedical Chemistry, University of
 Wisconsin, 1300 University Ave., Madison, WI 53706, USA
 JOURNAL
 FEATUES
 source
 1..6667
 location/Qualifiers
 /organism="Cloning vector pCAD-C1"
 /db_xref="taxon:52012"
 1..414
 /note="ADH1 promoter"
 488..829
 /gene="GAL4"
 <488..>829
 /gene="GAL4"
 /function="transcription factor"
 /note="transcription activation domain only"
 /codon_start=1
 /product="GAL4p"
 /protein_id="AB49958.1"
 /db_xref="GI:1595840"
 /translation="ANFNOSGNITADSSISTFTNSSNGPNLITTTQTSQALSDPIASS
 NVHDPNNEITRASKIDDDGNNKPLSPGWTDTAYNAFGITGMENTTDDVYNYLF
 DDEDTTPNPKE"
 833..874
 /note="polylinker region"
 1112..1441
 /note="ADH1 terminator"
 complement(1554..2648)
 /codon_start=1
 /product="LEU2"
 /protein_id="AB49959.1"
 /db_xref="GI:1595841"
 /translation="MSAPKKIIVLPDGHVGEITAEIKVLAISDVRSNVKFEDEHN
 LIGGAFAEATGVPDLEALASKADVAVLGAVGPKWGTGVRREOGILKTRKELOL
 VYLRPCNFAISDLSLDSPKIPQAFKGFVYVRELVGIIYFGKREDDGDVAMDE
 OYTVPEVORITRMAAFMAIOHEPPLPIWISLDKANVLAISRLRKTYEETIKNEPTLK
 VOHQLDSAAMLIVKNPHTLNSIITTSNMPGDIIDESVITGSGIILPSASLASLPD
 KNTAFGLVEPCHGSADLPKKNVDPITATILSAAMMLKLSLNPBCKAIEDXVKVYLD
 AGIRTGDLGSGNSTEVDGVAEVEVKILIA"
 3497
 /note="COLE1 replication origin"
 complement(4256..5116)
 /codon_start=1
 /product="Ampr"
 /protein_id="AB49960.1"
 /db_xref="GI:1595842"
 /translation="MSIOHFRVALIPFAFLPYFAHPETLVKVKDAEDQLGARVY
 IELDNGSKILESFRPEERPMWSTFKVLLCGAVISRDAGQELGRIRHSQNDLVE
 YSPVTEKHITDGMVRELCSAIIITMSDNTAAILTTIGSPKEFLIPLNMGDHVTRL
 DRWEPELNAIPNDERDITTPAAMATTLRLKLLIGELLTASRQQLIDWMEADKVAGPL
 LRSALPAGMFIADKSGAGERSRGIIAALGPGKPSRIYVITTSQATMDERNNOIA
 EIGASLIKHW"
 5506..6667
 /note="yeast 2 micron replication origin"
 BASE COUNT 1904 a 1501 c 1369 g 1893 t
 ORIGIN

Query Match 16.5%; Score 346.8; DB 14; Length 6667;
 Best Local Similarity 93.1%; Pred. No. 4.2e-71;
 Matches 363; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

[illegible]

RESULT	9		
CVU13188			
LOCUS	CVU13188	6650 bp	DNA -circular SYN
DEFINITION	Cloning vector pCAD10, complete sequence.		
Accession	U13188		
Release Date	29-MAR-1996		

SOURCE	Cloning vector pGAD10.
ORGANISM	Cloning vector pGAD10
REFERENCE	artificial sequence; vectors 1 (bases 1 to 6650)

TITLE	CLONTECH Vectors On Disc version 1.3
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 6650)

TITLE Direct Submission
JOURNAL Submitted (08-AUG-1994) Paul A. Kitts, CLONTECH Laboratories, Inc.

COMMENT

call (415) 424-8222 or (800) 662-2566, extension 1. For technical customers, please contact your local distributor. For technical information, call (415) 424-8222 or (800) 662-2566, extension 3. This sequence has been compiled from information in the sequence databases, published literature and other sources, together with partial sequences obtained by CLONTECH; this vector has not been completely sequenced. If you suspect there is an error in this sequence, please contact CLONTECH's Technical Service Department at (415) 424-8222 or (800) 662-2566, extension 3 or E-mail TECH@CLONTECH.COM.

FEATURES	Location/Qualifiers
SOURCE	1. .6650
	/organism="Cloning vector pGAD10"
	/db_xref="taxon:36574"
BASE COUNT	1901 a 1496 c 1362 g 1891 t
ORIGIN	

Query Match	16.5%;	Score 346.6;	DB 14;	Length 6650;
Best Local Similarity	96.2%;	Pred. No. 4.7e-71;		
Matches 355;	Conservative	0;	Mismatches 14;	Indels 0;
			Gaps	0;

[illegible]

RESULT	10
CYU07647	
LOCUS	CVU07647 6659 bp DNA circular SYN
DEFINITION	Cloning vector pCAD24, complete sequence.
Accession	U08745
	29-JAN-1997

SOURCE	Cloning vector pGAD424.
ORGANISM	Cloning vector pGAD424
REFERENCE	artificial sequence; vectors 1 (bases 1 to 6650)

AUTHORS Bartel, P.L., Chien, C., Sternberg, R., and Fields, S.
TITLE Using the two-hybrid system to detect protein-protein interactions
JOURNAL (in) Hartley, D. (Ed.),
CELLULAR IMPRINTS IN DEVELOPMENT: A PRACTICAL APPROACH

153-179;

REFERENCE 2 (bases 1 to 6559)

TITLE CLONTECH Vectors On Disc version 1.2
JOURNAL

REFERENCE	3 (bases 1 to 6659)
ATTMORP	W:++c D A

JOURNAL TITLE
Direct Submission
Submitted (11-MAR-1994) Paul A. Kitts, CIONTECH Laboratories, Inc.,
4030 Eshelman Way, Palo Alto, CA 94302 USA

COMMENT

424-8222 or (800) 662-2566, extension 1. International customers, please contact your local distributor. For technical information, call (415) 424-8222 or (800) 662-2566, extension 3. This sequence has been compiled from information in the sequence databases, published literature and other sources, together with partial sequences obtained by CLONTECH; this vector has not been completely sequenced. If you suspect there is an error in this sequence, please contact CLONTECH's Technical Service Department at (415) 424-8222 or (800) 662-2566, extension 3 or E-mail CLONTECH@CLONTECHNET.COM.

FEATURES	location/Qualifiers
source	1..6659
	/organism="Cloning vector pGAD424"
	/plasmid="n"
	/db_xref="taxon:33778"
BASE COUNT	1502 a 1499 c 1367 g 1891 t

LOCUS CVU70026 6666 bp DNA SYN 21-MAR-1997
 DEFINITION Cloning vector pGAD-C3 complete sequence.
 ACCESSION U70026
 VERSION U70026.1 GI:1595847
 KEYWORDS
 SOURCE Cloning vector pGAD-C3.
 ORGANISM Cloning vector pGAD-C3.
 REFERENCE 1 (bases 1 to 6666)
 AUTHORS James, P., Halladay, J., and Craig, E.A.
 TITLE two-hybrid selection in yeast.
 JOURNAL Genetics 144 (4), 1425-1436 (1996)
 MEDLINE 97132579
 REFERENCE 2 (bases 1 to 6666)
 AUTHORS James, P., Halladay, J., and Craig, E.A.
 TITLE Direct Submission
 JOURNAL Submitted (09-SEP-1996) Biomolecular Chemistry, University of Wisconsin, 1300 University Ave., Madison, WI 53706, USA
 FEATURES
 source
 1. 6666
 /organism="Cloning vector pGAD-C3"
 /db_xref="taxon:52014"
 promoter
 1. 414
 /note="ADH1 promoter"
 488. 829
 /gene="GAL4"
 <488. >829
 /gene="GAL4"
 /function="transcription factor"
 /note="transcription activation domain only"
 /codon_start=1
 /product="GAL4"
 /protein_id="AAB49964.1"
 /db_xref="GI:1595848"
 /translation="ANFNOSGNADSLSTFTNNSNGPVLITQNSQLSOPIAS
 NVHDFMNEITASKRIDGNSKPSLPGMTDQAYNAFGITTMFTTMDVYNLYE
 DDEDTPNPKE"
 833. 873
 /note="polylinker region"
 111. 1440
 /product="ADH1 terminator"
 complement(1553. 2647)
 /codon_start=1
 /product="Leu2"
 /protein_id="AAB49965.1"
 /db_xref="GI:1595849"
 /translation="MSAPKKIVLPDGHVQGETTAETAKVLSIDYRSVKEDFENH
 LIGGAETATGVPFLPDEALBASKVDVAVLGAAGVPRWGVSVPREGGLKIRKEDL
 VNLPRCNASDLSLDLSPKQFAKGTDFVVRVGLVGYFRKREDDGDAWQSE
 QYTVPEVQRTTRMAAFMALQHEPPLIWSLDKAVNLASSRLMKRTVEETIKNEPTLK
 VOHOLDSAMITLVKNPTLNGLIITSNMGGDITSPASVIPSGLGLPASIASLIPD
 KMAFGLYEPCHGSADPLPKNKVDPIATITLSAAMKLKSLNLPBEGKAIDAVKVIAD
 AGRIGDDLGSGNSTTEVGDAVAEEVKIITA"
 3496
 /note="ColEI replication origin"
 complement(4255. 5115)
 /codon_start=1
 /product="AmpR"
 /protein_id="AAB49966.1"
 /db_xref="GI:1595850"
 /translation="MSIQHFRVALIPEFAFCULPVFAHPETLVKVKDAEDQLGARVGY
 IELDLSGKILLESFRRPEFRPMSTFKVLLCGAVLSRIDAGQOLGRIRHYSONDLVE
 YSPVTEKLHLDGTVRELCSAAITMSDNTAANLLLTIGSPKELTAFLHMGDHVRL
 DHWEPLKEALIPDEEDRTTMVPVMAATPTIKRLTGLLTLASROQLIDMMADKVAAGL
 LRSALPAGNFIADKSCASGERGSGITIAALGPDSKPSRTIVITTTGSOATMDERNROLA
 EIGASLIKHM"
 5505
 /note="yeast 2 micron replication origin"
 rep_origin
 1904 a 1500 c 1369 g 1893 t

Query Match 16.5%; Score 346.6; DB 14; Length 6666;
 Best Local Similarity 96.2%; Pred. No. 4.7e-71;
 Matches 355; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
 QY 50 aagaagaagcggaaggtctcgagcgcgccaatttaacaaagtgggaattctgac 109
 |||||
 Db 461 AAGAGAAAGGTCGATTTGGGTACCCGCCCAATTTTAAACAAAGTGGGAAATTCCTGAT 520
 QY 110 agcatatgcttcctacttcctacacagtaacagtcgcgaacttaacaaactaa 169
 |||||
 Db 521 ACGTCATGTCCTTCCTTACCTTTCCTACATACAGTACCAACGCTCCCACTCTTAACAACTCA 580
 QY 170 acaattctcaagcgcttcacaaacaaattgctctcctcctcaacgctacatgatactcaltg 229
 |||||
 Db 581 ACAATTCCTCAAGCGCTTTCACAAACCAATTCCTCCTCTTAACGTTATGATTAATCTCAG 640
 QY 220 aataatgaatcagcgctcagtaaaattatgattgtaataatcaaacaccatgcaact 289
 |||||
 Db 641 AATAATGAATTCACGCGCTAGTAAATTTGATGATGATTAATTAACAAACCACTGTCACCT 700
 QY 290 ggttgacgagccaaactgcgtatagcgcttlttgatcaatcaacagggatgtttaacc 349
 |||||
 Db 701 GGTTCGACGCGACCAACTGCGTATACCGGCTTTGGAATCCTACAGGGATGTTTATACCT 760
 QY 350 actacaatgagatgatalataactatcattcgaatgataagatacccaaccaccca 409
 |||||
 Db 761 ACTACAATGAGATGATGTATATATCATCTATTCGATGATGAAGATACCCACCAACCA 820
 QY 410 aaaaagaag 418
 |||||
 Db 821 AAAAAAGAG 829

RESULT 13

AF134167 7389 bp DNA SYN 30-DEC-1999
 LOCUS Yeast 2-hybrid vector PC-ACT.1 complete sequence.
 DEFINITION AF134167
 ACCESSION AF134167
 VERSION AF134167.1 GI:6642848
 KEYWORDS
 SOURCE Yeast 2-hybrid vector PC-ACT.1.
 ORGANISM Yeast 2-hybrid vector PC-ACT.1.
 REFERENCE 1 (bases 1 to 7389)
 AUTHORS Durfee, T., Draper, O., Zupan, J., Conklin, D.S., and Zambryski, P.C.
 TITLE New tools for protein linkage mapping and general two-hybrid screening
 JOURNAL Yeast 15 (16), 1761-1768 (1999)
 MEDLINE 20060013
 REFERENCE 2 (bases 1 to 7389)
 AUTHORS Durfee, T., Draper, O., and Zupan, J.
 TITLE Direct Submission
 JOURNAL Submitted (11-MAR-1999) Plant and Microbial Biology, UC Berkeley, 111 Koshland Hall, Berkeley, CA 94720, USA
 FEATURES
 source
 1. 7389
 /organism="Yeast 2-hybrid vector PC-ACT.1"
 /db_xref="taxon:111295"
 1. 362
 /note="ColEI replication origin cloned from pRS315;
 SIKORSKI, RS and HIELETER, P. 1989"
 complement(934. 1794)
 /gene="bla"
 complement(934. 1794)
 /gene="bla"
 /note="confers ampicillin resistance; cloned from pRS315;
 SIKORSKI, RS and HIELETER, P. 1989"
 /codon_start=1
 /product="beta-lactamase"
 /protein_id="AAP20324.1"
 /db_xref="GI:6642849"
 /translation="MSIQHFRVALIPEFAFCULPVFAHPETLVKVKDAEDQLGARVGY
 IELDLSGKILLESFRRPEFRPMSTFKVLLCGAVLSRIDAGQOLGRIRHYSONDLVE

[illegible]

OY	290	gattgcgcgcgcacccaactcgtcgtataacgcgttttgaaatcaatcaacgaagatgatttaacc	349
Db	6662	GGTTGGAGCGACCAACTGCGTATACGCGGTTTGGAATCTACTACAGGGAATGTTTAATACC	6721
OY	350	actcaatcagatgattgatatataactatcattcgtatgatgaagaatacccaaccca	409
Db	6722	ACTACAAAGATGATGTATATATAACTATCTATTCGATGATGAATACCCACCAACCA	6781
OY	410	aaaaaagg	418
Db	6782	AAAAAAGAG	6790
RESULT 14			
AF134168			
LOCUS	AF134168	7404 bp	DNA
DEFINITION	Yeast 2-hybrid vector pc-ACT.2 complete sequence.		30-DEC-1999
ACCESSION	AF134168		
VERSION	AF134168.1	GI:6642852	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	1 (bases 1 to 7404)		
TITLE	Durfee,T., Draper,O., Zupan,J., Conklin,D.S. and Zambryski,P.C.		
JOURNAL	New tools for protein linkage mapping and general two-hybrid		
MEDLINE	screening		
REFERENCE	Yeast 15 (16), 1761-1768 (1999)		
AUTHORS	2 (bases 1 to 7404)		
TITLE	Durfee,T., Draper,O. and Zupan,J.		
JOURNAL	Direct Submission		
MEDLINE	Submitted (11-MAR-1999) Plant and Microbial Biology, UC Berkeley,		
REFERENCE	111 Koshland Hall, Berkeley, CA 94720, USA		
AUTHORS	Location/Qualifiers		
TITLE	1..7404		
JOURNAL	/organism="Yeast 2-hybrid vector pc-ACT.2"		
MEDLINE	/db_xref="taxon:111296"		
REFERENCE	1..362		
AUTHORS	/note="COLE1 replication origin cloned from pRS315;		
TITLE	Sikorski, RS and Hieter, P. 1989"		
JOURNAL	complement(934..1794)		
MEDLINE	/gene="bla"		
REFERENCE	complement(934..1794)		
AUTHORS	/gene="bla"		
TITLE	complement(934..1794)		
JOURNAL	/gene="bla"		
MEDLINE	/note="confers ampicillin resistance; cloned from pRS315;		
REFERENCE	Sikorski, RS and Hieter, P. 1989"		
AUTHORS	/codon_start=1		
TITLE	/product="beta-lactamase"		
JOURNAL	/protein_id="AAF20327.1"		
MEDLINE	/db_xref="GI:6642853"		
REFERENCE	/translation="MSIOHFRVALILPEPAACFLPFAHPETIVKVKAEADOLGARVGY		
AUTHORS	TELDNSKILIESFRPERFERPFMASTFRVLCAGVLSRIDAGQEOIGSRHYSQNDIYEL		
TITLE	YSPVTEKLIDVGMTVRELCSAATITMSDNTANLLITITGGKELTFARFHHNGDVTYEL		
JOURNAL	DRWPELNEALPNDERDTMPVAMATTLRLKLTGELLTLASROOLDIMWEADKVGFL		
MEDLINE	LRSAIPAGMFTADKSGAGERSGIIAALGDPKPSRIVYITTSQATMDERNROIA		
REFERENCE	EIGASLIKHW"		
AUTHORS	2028..2053		
TITLE	/note="CEN6: CDE1 fragment of chromosome VI centromere		
JOURNAL	from S.cerevisiae; cloned from pRS315; Sikorski, RS and		
MEDLINE	Hieter, P. 1989"		
REFERENCE	2066..2442		
AUTHORS	/note="ARS4: Autonomous replicating sequence fragment		
TITLE	near gene for histone 4 on chromosome II in S. cerevisiae;		
JOURNAL	cloned from pRS314; Sikorski, RS and Hieter, P. 1989"		
MEDLINE	complement(3171..4265)		
REFERENCE	/gene="LEU2"		
AUTHORS	complement(3171..4265)		
TITLE	/gene="LEU2"		
JOURNAL	complement(3171..4265)		
MEDLINE	/note="cloned from pRS315; Sikorski, RS and Hieter, P.		
REFERENCE	1989"		
AUTHORS	/codon_start=1		